

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 14, 2001, 16:40:22 ; Search time 7696.04 Seconds
(without alignments)
4464.708 Million cell updates/sec

Title: US-09-301-906-3
Perfect score: 6714
Sequence: 1 atggactacattcgccatt.....gttgcttcgaggatatga 6714

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1118133 seqs, 2558875100 residues

Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba1.*
- 2: gb_ba2.*
- 3: gb_om.*
- 4: gb_ov.*
- 5: gb_ph.*
- 6: gb_pl1.*
- 7: gb_pl2.*
- 8: gb_pri.*
- 9: gb_pr2.*
- 10: gb_pr3.*
- 11: gb_fun.*
- 12: em_hum1.*
- 13: em_hum2.*
- 14: em_in.*
- 15: em_om.*
- 16: em_or.*
- 17: em_ov.*
- 18: em_pat.*
- 19: em_ph.*
- 20: em_pi.*
- 21: em_ro.*
- 22: em_sts.*
- 23: em_sy.*
- 24: em_un.*
- 25: em_v1.*
- 26: gb_htg1.*
- 27: gb_htg2.*
- 28: gb_in1.*
- 29: gb_in2.*
- 30: em_ba1.*
- 31: em_ba2.*
- 32: em_hum3.*
- 33: em_hum4.*
- 34: gb_pr4.*
- 35: gb_htg3.*
- 36: gb_htg4.*
- 37: gb_htg5.*
- 38: gb_htg6.*
- 39: gb_htg7.*
- 40: em_htg1.*
- 41: em_htg2.*
- 42: em_htg3.*
- 43: em_hum5.*

- 44: gb_pl3.*
- 45: gb_pr5.*
- 46: gb_htg8.*
- 47: gb_htg9.*
- 48: gb_htg10.*
- 49: gb_htg11.*
- 50: gb_htg12.*
- 51: gb_htg13.*
- 52: gb_htg14.*
- 53: gb_in3.*
- 54: gb_htg15.*
- 55: gb_htg16.*
- 56: gb_htg17.*
- 57: em_htg4.*
- 58: em_htg5.*
- 59: em_htg6.*
- 60: em_htg7.*
- 61: em_hum6.*
- 62: gb_htg18.*
- 63: gb_htg19.*
- 64: gb_ba3.*
- 65: em_htg8.*
- 66: em_htg9.*
- 67: em_htg10.*
- 68: gb_pr6.*
- 69: gb_pr7.*
- 70: gb_htg20.*
- 71: gb_htg21.*
- 72: gb_htg22.*
- 73: gb_htg23.*
- 74: gb_ro.*
- 75: gb_sts1.*
- 76: gb_sts2.*
- 77: gb_sy.*
- 78: gb_un.*
- 79: gb_v1.*
- 80: gb_v12.*
- 81: gb_pat1.*
- 82: gb_pat2.*
- 83: em_htg0.*
- 84: gb_htg24.*
- 85: gb_pr8.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2557	38.1	4173	81	AR070579 Sequence
2	2107	31.4	13154	79	AF037268 Grapevine
3	358.6	5.3	14861	79	AF283103 Pineapple
4	220.2	3.3	12394	79	AF195822 Grapevine
5	87.4	1.3	7349	79	X75433 Grapevine v
6	80.8	1.2	9306	79	D21829 Apple stem
7	75	1.1	6790	79	AF034622 Tobacco r
8	75	1.1	6791	79	AF166084 Tobacco r
9	75	1.1	6791	80	D00155 Tobacco rat
10	75	1.1	6791	80	X06172 Tobacco rat
11	73.4	1.1	1472	80	Z97358 Tobacco rat
12	71.8	1.1	6828	80	L23972 Pepper ring
13	70	1.0	6656	80	D13957 Papaya mosa
14	69.8	1.0	9293	79	AB045371 Apple ste
15	69.6	1.0	7073	80	X14006 Pea early b
16	63.4	0.9	10545	79	BY051931 Beet yellow
17	62	0.9	8832	80	M97264 Shallot vir
18	59.6	0.9	8118	80	U15440 Lettuce inf
19	57.6	0.9	4683	80	U47034 Sunn-hemp m
20	57.4	0.9	7599	79	X75448 Grapevine v
21	57	0.8	19259	79	AF260651 Citrus tr

QY 3922 gtgcgcagctgggattagcaccacaggggaagttaacttggtgaagaaagagaccattca 3981
Db 1381 GCGGTGGCTTCGACCGCGGTGAGGCAGGTCCAGCCATGCTGTTGAATCACCAC 1440
QY 3982 cctcgaagcgtgcaatcacacttatgtcggaggaaggttgcccgctctcagcgtggcg 4041
Db 1441 AGGGGTAGTGTGAGTTCGCTACAGAACGCTGTAGATGAACATCCGCGGTGTGTA 1500
QY 4042 gaaagacaaggtgatccgtcggtgttctgtgtaaccgtgacgtatgtgtttttgaaagt 4101
Db 1501 GCTGAAGTTTGAATAA-----GGATCTAAATAACACCACCTTGGTACAGCTGCTCTAGAGTCG 1554
QY 4102 gtgaaaaaaggtgtcgacgatctttccaccaacagctctagtggggaaacggtcgtgag 4161
Db 1555 CCCCCGTAGGTCTGAAGCTGGAGCGCGCCCAACGCTCAGAGACGGTGTGTCGGAGGTT 1614
QY 4162 gttgaggtggaagcgcaaggttgctcccgagaaagcgtcgtcggtgagcgccgacacaa 4221
Db 1615 GAAGCTGAGAANTGTCGGAGGTCATCGTTGACGTTCTAGTTCAGAACCGCGGTACAA 1674
QY 4222 gaaaggggaagcgtgcagatgttaacacagcacaacacccggtgtaacagagcgacagg 4281
Db 1675 GAAGTCTTGAATCAACCAATGCTGCCAAGCTGCAAGAACCTGAAGAGTGTGCGAGGCG 1734
QY 4282 gagccagttacagttccagtcctgtgagttcgccacaggtctgatatctccaaaggtcaccceag 4341
Db 1735 GACACATGTGGAGCTGGGGTAGCTAATCAGAACTGAGTCAACGTTGTTTCTCGCGCAA 1794
QY 4342 tccgaggtacat---gctcagaagaagtgaacaagagtagtaccattggcgactgtttcg 4398
Db 1795 GTACCCGCACATGAGCTGGTCTTGAGGCATCTAGTGGCGCGGTCGTGAGGCCATTGCCAA 1854
QY 4399 ggcgcacccaaatcgtctgatgagaacccgccccaaagtgt---tacgactcgtggtgtg 4455
Db 1855 GTTTCTGTGCCAGTAGCCGTAGAGAAAACCTGTTTATCTGTGAGAAGCGCGGTGAGCTA 1914
QY 4456 aagataattgacaaggaagccgtcgtctatgtgctgagaaaaaaacaggtacaagtc 4515
Db 1915 AAGCGGTAGATAAGGCAAGCGGTGTCGACGCAAGAAAGTCAAGAAATATACCGGTT 1974
QY 4516 gagcagcccaacagagaggtttgacgatcaatgaagcgaagcggttaaacagctttgc 4575
Db 1975 AAGCGTTACACGAGGGGCTCTAAAAATTAGTAGGATACCGTTCTGAAGNAATTGTGC 2034
QY 4576 atgtttagaacgtgttcctcgtggtgtgcagctggatgtgtacaagaaacgactatcgcc 4635
Db 2035 ATGTTTAGAACGTGTCTCGCGCGCTGCGAGTTGGACGTGTACAAATGAAGCGACCATCGCC 2094
QY 4636 accaggttctcaaacgcatctacccttgcgataactgaagggagaggtcggtcttt 4695
Db 2095 ACTAGGTTCTCAATCGGTTTACCTTTGTGATAGCTTGAAGGGAGGAGTGCGGTCTTT 2154
QY 4696 ttctcaagctgggtgaggggtatacctataatggtgtgtagccatgtttcatcaggggtg 4755
Db 2155 TTCTCAAGCTGGGTGAGGGGTATACCTATANTGTTGTTAGCCATGTTTCATCAGGGTGG 2214
QY 4756 cctcgtccctgagagatatcttaacggcaataagttaccacgaagcgtcttcgaccactgt 4815
Db 2215 CCTCGTGCCTAGAGGATATCTTAACGGCAATTAAAGTACCCCAAGCGTCTTCGACCACCTGT 2274
QY 4816 ttatgcagaagtacaagtgggtgagggcgtaccattccacgcgtgatacaggaagtc 4875
Db 2275 TTAGTGCAGAAGTACAAAGTGGTGGAGCGGTACCATTCACGCTGATCAGGAGGAGTGC 2334
QY 4876 tatccatcagataaacctatcttgacggttcaatctcgtggggaagcaactcttcgact 4935
Db 2335 TATCCATCAGATAACCTATCTTGACGGTCAATCTCGTGGGAGAGCAAACTCTCGACT 2394
QY 4936 aagtcgaggaaggtgtgtaagtgatgtgtcataaacgtagcttccggtgactattttct 4995
Db 2395 AAGTCGAGGAAGGGTGGTAAGGTGATGTCATAAACGTAAGCTTCGGGTGACTATTTTCTT 2454
QY 4996 atgccttcggtgttttcaaaaggacgcacttgcatctcagtaaacctccatcgacgaagggcg 5055

Db 2455 ATGCCCTGCGGGTTTCAAAGGAGGCACCTTGCAATTCAAGTAAATCCATCCACGAAGGCGC 2514
QY 5056 atcagtttgacgttccagggcaactcggcgctcttgggtgtagcgaggtggttcgagttca 5115
Db 2515 ATCAGTTTACAGCTTACAGGCAACTCGCGCGCTTTGGTGTAGCAGGATGTTGCAGTTA 2574
QY 5116 gtcggcggtgctcggtgaggaagtcaccaggtgttccaaaccagcaacacacagagccaa 5175
Db 2575 GCGCGCGGCTGTGCGATGAGAAGTCAACAGGTGTTCAAACACAGCAACACACAGAGCAA 2634
QY 5176 gttgctaccagaaacaateacacaaaaatcggggggcaaggtctctatctgaggaagtggt 5235
Db 2635 GGTGCTTACCAGAACAAATCACCCAAATCGGGGGCAAGGCTCTATCTCAGGGAAGTGGT 2694
QY 5236 agggaaagtcgaagggagtcgacatactcgatatggtgcgaacaaagatcacgttaggaag 5295
Db 2695 AGGGAAGTCANAGGGAGGTGCGACATCTCGATATGTTGGTGAACAAAGTTACGTTAGGAAG 2754
QY 5296 tgtgagtggtcagggctgataatccagtgatgggtcttgaacctgactacaccccaatg 5355
Db 2755 TGTGAGTGGCTCAGGGCTGATAATCCAGTGTGGCTCTTAAACCTGGCTACACCCCAATG 2814
QY 5356 acatttgaagtggttaaaaccgggaacctctgaagatgcgctcggtgaggtacttgaagtat 5415
Db 2815 ACATTTGAAGTGGTTAAAGCCGGGACCTCTGAAGATGCGCGTCTGTGAGTACTTTGAAGTAT 2874
QY 5416 ctggctataaggtcattgagaggacatacagggcggtgctctatgcttagaataatgcgctc 5475
Db 2875 CTGGCTATAGGCATTTGGGAGGACATACAGGGCGTTCCTTATGGCTAGAAATATTTGCCGTC 2934
QY 5476 actaccgcgaaggtgtctgaaagtacctaatacaagtttatgaatcaactacccggcttt 5535
Db 2935 ACTACCGCGGAAGGTCTCTGAAAGTACCTAATCAAGTTTATGAATCACTACCCGGCTTT 2994
QY 5536 cagtttacaagtcgggcagagatctcattttttcatctcaacaaagagcggcttcggtg 5595
Db 2995 CACGTTTCAAGTCGGGCGACAGATCTCATTTTTCATTCAACACAGACGGCTTTCGCTGTG 3054
QY 5596 agagacctaccgtacgtactactagctgaaaaggtatctttaccagagggcaagatgtc 5655
Db 3055 AGAGACCTACCGTACGTATTCATAGCTGAGAAAGGTATTTTATCAANGGCAAGATGTC 3114
QY 5656 gacgoggtggttagcttggggcacaatctgttcgtatgcgacgataactggttttccac 5715
Db 3115 GACGCGGTAGTAGCTTTGGCGCACAATCTGCGTATGTCATGATATATTGTTTCCAT 3174
QY 5716 gatccatttaattgaataggtgcactgaagtcgctcgatgcggaatggtggcggaatcg 5775
Db 3175 GATGCTATTAAATTTGATGGGTGCACCTGAAGTTGCTCGATGTTGTTGGTGGTGAATCA 3234
QY 5776 tttaagctccttcgaaataaagtgtataatgctccccaggtggtcggttaagacgcagcg 5835
Db 3235 TTTAAGTCGTTCGAATACAAATGCTATANTGCTCCCCAGGTGCGGTGAAGACGAGATG 3294
QY 5836 ttagtggaagaaatctgtaagtcacccaatagcagacccaattacgctgaatgtggga 5895
Db 3295 CTAGTGGAGCAATTTGTCAAGTCACCAATAGCACGCCACCAATTACGGCTAACGTGGGA 3354
QY 5896 agttctgagacataaataatgcggtgaaagagagatccgaatttgaaggtctcaac 5955
Db 3355 AGTTCTGAGGACATAAATATGCGGTGAAGAGAGAGATCCGGAATTTGGAAGGCTCTCAAC 3414
QY 5956 agtgcctaccacaggttaactccaggtgtglaaactttatcgtcaggggaatgtataaagg 6015
Db 3415 AGTGTACACAGTTAACTCCAGGTGGTGAATTTATTTATGTCAGGGGAATGTATAAAGG 3474
QY 6016 gttttggtggaatgaggtgcacatgactcaaggtctactacaactagcgtcttcgca 6075
Db 3475 GTTTTGGTGAGGTGATGATGATCAAGGCTTACTACAACCTAGGCGTCTTCGCA 3534
QY 6076 accggcggtcggaagggcctcttttttgagacataaaacagataccattcataaaggg 6135

Db	3535	ACCGGGCGCTCGGAAGGCGCTCTTTTGGAGACATAAATACAGATACCATTATATAACCGG	3594	JOURNAL	Submitted (08-DEC-1997) Plant Pathology, Cornell University, New York State Agricultural Experiment Station, Geneva, NY 14456, USA
Qy	6136	ggaagggttaggatgattgctgtgttttgcctccaaagaagaagcgtttgtatatac	6195	COMMENT	On Apr 29, 1998 this sequence version replaced gi:23351133.
Db	3595	GAGAAGCTTTAGGATGGATTGCTGTTGTTGTTCCAAAGAGAAAGCGTTGTATAC	3654	FEATURES	Location/Qualifiers
Qy	6196	acttctaatactgacaggtgctgttagatgtttgctactgtgttcctcaatgaccgta	6255	source	1..13154
Db	3655	ACTTCTAAATCATACAGGTGTCGGTTAGATGTTGCTACTTGTTCCTCAATGACCGTA	3714		/organism="grapevine leafroll-associated virus 3"
Qy	6256	aggggaaggaagaadtgtaccctgaagaagtcgttagcgtgaagacaacaaccagtagta	6315	CDS	/strain="NY1"
Db	3715	AGGGGAACGGAAGTGTACCCTGAAAAGTCGTAGCGGTGAAGCAAAACCCAGTAGTA	3774		/db_xref="taxon:55951"
Qy	6316	agatcgctgtccaaaggccaaattggaaccactgatgcgtagctgaaataaaacgctgac	6375		<1..2107
Db	3775	AGATCGCTGTCCAAAGGCCAATTCGAACCACTGATGACGTAGCTGAATAAACCGCTGAC	3834		/function="replication and unknown functions"
Qy	6376	glgtactgtgcatgacccagttgagaagtcggatatgaagaggtcgttgaaagggaataa	6435		/note="incomplete ORF1a, one-third of the expected ORF1a; sequenced portion contains helicase domains identified by sequence comparison"
Db	3835	GTGTACTTGTGCATCACCCAGTTGGAGAAGTCGGATATGAAGAGGTCGTTGAAGGCAAAA	3894		/codon_start=2
Qy	6436	ggaagaagaacacacagtgatgacatgcatgaagcacaggggaacaaacattcagtgatgtg	6495		/product="unknown"
Db	3895	GGAAAGAAGAACACCACTGATGATGATGATGAAGCACAGGGAACAACTTCAGTGATGTG	3954		/protein_id="AAC40717.1"
Qy	6496	gtatgttttagacgaagaagccgatgactccctattcactaaacacccgcataatactt	6555		/db_xref="GI:3090885"
Db	3955	GTATTGTTTAGACCAAGAACCCGATGACTCCCTATTTCACTTAACACCCGATATACTT	4014	CDS	/translation="DYNEATATFRSNAFTFDNLKRSVAFVFSKLGEGYTYGGSH VSSGPRALDILTAIRKPYVSFVDFCLVQYKMGGVFPFADDEECYFSDNPILVNLV GKANFSTCKRGGKVMVINVASGDYFLMPCGFQTHLHSVNSIDEGRISLTFRATRV FGVRMLQLAGVSDKSPGVPNOQPOSGATRTITPKSGGKALSEGSGREVKGRTV STWCEDYVRKCEMLRADNPVMALEPDYTPMTEFVVKTCGSEDADVBYLKLALGIER TYRALLMARNIAVTAGVLKVPNQVYESLPGFHVYKSGTDLIIFHSTQDGLURVADLPY VLIAEKGIFTKGDDVAVDALGNDLFCDDILVFDHDAINLIGALKVACRGKMGVSFYS FEYKYNAPGGGKTTLLVDFEVSFSPNSTATITANVGSSDINMAVKRDPNLEGLNS ATTNSRVNFIVRGMYKRVLDVDEHMHQGLQLQGVFATGASEGLFGDINOIPFN REKVERMDCAVFVPKESVYTSKSRCPDVCYLLSSMTVRGTEKYPEKVVSGKDK PVVSLSKRPIGTTDDVAEINADVYLCMTOLEKSDMKRSLKGCKETPVMPVHQAQK TFSVDLVFRYTKKADDSLTFTQPHILVGLSRHTRSLVYAALSSKLDKDKVGYTISDASPO SVSDALLHTFAPAGCFRGI"
Qy	6556	gttgtttgtgagacacacacgctcactggtttatgcgcgtctgagctcaagttggac	6615		/function="replication"
Db	4015	GTGTGTTTGTGAGACACACACACGCTCACTGGTATTATGCGGCTCTGAGCTCAGAGTTGGAC	4074		/note="ORF1b; probably expressed via +1 frameshift mechanism; 61 kDa RNA-dependent RNA polymerase"
Qy	6616	gataaggtcggcacatatattagcagcgcgtcacctcaatcagtcagagcgtttgctt	6675		/codon_start=1
Db	4075	GATAAGGTGCGGCACATATATTAGCAGCGCTGCGCTCAATAGTAGTCCGACGCTTGCTT	4134		/product="RNA-dependent RNA polymerase"
Qy	6676	cacagcttcgcccgcgtggtgcttcgaggtatatga	6714		/protein_id="AAC40705.1"
Db	4135	CACAGTTCGCCCCGCGCTGGTGTGCTTCGAGGTATATGA	4173		/db_xref="GI:3090873"
RESULT	2				/translation="MNPCTPEGELVRKIPTSHEFVAVNGFLEDDLDGCPAFDYDFED DFETSDOSFLIEDVRISESFSHTSKIEDREYFIRSSVGI.PKRNLTLCNLTAVTENRN FNADRGCVGDDSDVAHELKEITFEFEEVYNKARLAETPSHLSSTNTMLLSMDLDRAPN AYSLKRALGDSVFHPHSMLTSTYLLMADVKNRKLNDTPLSKYVTGQNIIVYHDCRVATL FSCIFTACVERLYVDERWFLFHGMDTAEALAAALRNRLNDIRQIYTYELDISKYDKS QSALMKQVEELILTLTGVDRVELSTFFCGEYDSVVRTMTELVSQSRSGSGANTW LGNSLVLCTLLSVYLRGLDYSYIVVSGDSDLSIFSRQPLDIDTSLVSNFGDFDKIFNQ AAFYPCSKFLVQVEDSLFFVPDPLKLVKFGASKTSDIDLLEHIFQSFVDLSKGFNRE DVIGELAKLVTRYKYKHSGWTYSALCVLHLVLSANFSQPCRLYYHNSVNLVDVRIQRTES LSLALAKARILURWKRASFASIKRG"
LOCUS	AF037268	13154 bp	RNA		/function="unknown"
DEFINITION	Grapevine leafroll-associated virus 3, partial genomic sequence.				/note="ORF2; seems to have no homologs in genomes of other closteroviruses"
ACCESSION	AF037268	U82337			/codon_start=1
VERSION	AF037268.1	GI:3090872			/product="6 kDa protein"
KEYWORDS	grapevine leafroll-associated virus 3.				/protein_id="AAC40706.1"
SOURCE	grapevine leafroll-associated virus 3				/db_xref="GI:3090874"
ORGANISM	Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae; Closterovirus.				/translation="MYSRGSPFKSRVLTPLTVLGAYMWEFELPYLTDRKRHSYSAPSVA TFSLVSR"
REFERENCE	1 (bases 7833 to 9860)				5165..5302
AUTHORS	Ling, K.S., Zhu, H.Y., Alivizo, H., Hu, J.S., Drong, R.F., Slightom, J.L. and Gonsalves, D.				/function="unknown"
TITLE	The coat protein gene of grapevine leafroll associated closterovirus-3: cloning, nucleotide sequencing and expression in transgenic plants				/note="ORF3; similar to small hydrophobic proteins of other closteroviruses"
JOURNAL	Arch. Virol. 142 (6), 1101-1116 (1997)				/codon_start=1
MEDLINE	97372946				/product="5 kDa protein"
REFERENCE	2 (bases 1 to 13154)				/protein_id="AAC40707.1"
AUTHORS	Ling, K.S., Zhu, H.Y., Drong, R.F., Slightom, J.L., McFerson, J.R. and Gonsalves, D.				/db_xref="GI:3090875"
TITLE	Nucleotide sequence of the 3'-terminal two-thirds of the grapevine leafroll-associated virus-3 genome reveals a typical monopartite closterovirus				/translation="MDDFKQAILLLVDFEVIILLVLTFFVPRPQQSSPTINTGLRT V"
JOURNAL	J. Gen. Virol. 79 (Pt 5), 1299-1307 (1998)				5321..6970
MEDLINE	98264508				/function="molecular chaperone"
REFERENCE	3 (bases 1 to 13154)				/note="ORF4; identified by sequence comparisons; similar to HSP70"
AUTHORS	Ling, K.-S.				/codon_start=1
TITLE	Direct Submission				


```

/product="59 kDa protein"
/protein_id="AAC40708.1"
/db_xref="GI:3090876"
/translation="MEVGIDFGTFTSTICFSPSGVSGCTPVAGSVVETQIFPIEGSS
TVLIGAAKAYRDGVEGLYVNPWRAGVTRDNVRYVEKLPVTVTKIDSGGALLI
GGLGSPDVLVAVDVICLFLRALILECERYTSTVTAAVTVPADYNSFKRSEFVEA
LKLGLTPRGVNEPTAAALYSLAKSRVEDLLAVDFGGGTFDVSFVKKGNILCVI
FVSGNDLGGDRIDRAIVEIKKIKGKADAKLGLIFSSMKEDUSNNNAITQHLPIV
EGVEVVDLDELDAIVAFPSARAVEKTLGNFYPDPVIAVMTGSSALVKVRSD
VANLPQISVDFSDFRCSVACAKYCDTLAGNSGLRDLVTNLTI/DEVVGLQPV
VIFPKGSPICSTHYRTVGGGVVYVIGEGNNRAFLNEPTFRGVSKRKGDPVETDV
AOFNLSTGCTSVVNGVEVKVEYLVPGTTLNVLDSLVYKSGREDELEAKAIPEYLITLN
ILHDKAFTRNLGNKDKGFSDLRIENFLKSAVDVDTILNG"
6963..8414
/function="unknown"
/notes="ORF5; similar to genes found in a similar position
in other closterooviruses"
/codon_start=1
/product="55 kDa protein"
/protein_id="AAC40709.1"
/db_xref="GI:3090877"
/translation="MDKIYIVTCILNPNEARDEVSFVNKGYIGPGRSFSNRGSKYT
VWNSAARISGTSSTQSOTIDAFAYFLKGLGTTLSNPINCENWVRSSKDLSPAFR
TLIKGIYASRVSDNLPKKDRDDIMEASRRLLSPSDAAFCRAVSQVQGVYDVTONLE
STIVLRVMEIKRGSAHVSIPKVSAYDYTNLQELLSDEVTRARDTVSAYATD
SNAFLVMP/LTAREQWLADVLGILLVRRPANSYDVRVAVDYDIATLKLVLRF
NKDTGGIKNDLPCVPIESDFPFHELSSFSRLSEBMTTGGKICPCETAERLVRLLM
ENYKLR/LPVMALIIILVYISYIGTYNATRIKRRPDLNVRIRKGRVKS/LRGVEDRA
FRISKRGINAVRCLRYSDLTCLARRHYGIRRNWKTLSYVDGT/LAYDTADTICK
VNTINTADHASIIHVIKTNEQVGTTLPHQL"
8504..9445
/notes="ORF6"
/codon_start=1
/evidence=experimental
/product="35 kDa coat protein"
/protein_id="AAC40710.1"
/db_xref="GI:3090878"
/translation="MAFELKJGQIYEVVPPENNLVRVGDAAQKFSKASFLKYVDGT
QALGTIAVVPKPYFATAALAAQAPPPQPAQVAPQETDYGVSPESETLTPNKL
VEKDPDKFLTKMGKIALDLAGVTHKPKVINEPKVSVSEVAMKINAA/LME/LCKVNG
ADDAATKTEFFLYVMOIACFTTSSTEKFEFDYIETDGGKKIYAVWYDCKIQAAAS
TGVENPVROYLAQFPTEITATLNGKLVNKEKVMQAQGVPPKFFPYTIDCVRTYDLF
NNDAILANLWLRQAQRNKTVTADNTLHNVFQLLQKK"
9508..10941
/notes="ORF7; diverged copy of the GLRaV-3 coat protein"
/codon_start=1
/product="53 kDa protein"
/protein_id="AAC40711.1"
/db_xref="GI:3090879"
/translation="MGAYTHVDFHESRLKKQDYLSFKSADAEAPDPGYPVRPSYV
RAYLORADFPNTQSLVSLTSLASNKLSGLMSDAVSSFM/LMNDVGLYFECGVCHN
KPYLGRVIFCRKYIGRGVEITTKGNTYTSNNWNEASYVIQVNVVDGLAQTTVNSTYT
QTDVSG/LKNWTRIVYKTIIVSDQNLKPGCFSDSKLGVMRI/SLLSVPVRFRLIL
LAPLKKSFNARIEDVLNIDTSLVPSVPVPESTGCVGPGSEOLDVVALTSDVELINT
RQGKICFPDPSVLSNEADYDERVILPITEALQINARLRLVLSGGSGT/PRDMGMI
VAMIQDVLVYSTVKNLSKADGRVETELGQKRVYLSYSEVREALIGLGGYKYGASPTNVR
SPMRVFAHTITLIEKKIQPACTALAKHGVPKRPFPYCFD/FALLDNRYYPADVLKAN
AMACAIAIKSANLRKRGSETYNILESI"
10952..11509
/function="unknown"
/notes="ORF8; seems to have no homologs in other
closterooviruses"
/codon_start=1
/product="21 kDa protein"
/protein_id="AAC40712.1"
/db_xref="GI:3090880"
/translation="MEFRPVLTIVRRDPGVNTGSLKVIAYDLHYDNIFDNCVAKFSFRD
TDTGFTVMEYSTNAFSLPYKLFSAVENKEGEMISNDVSGSFRYINIFSQMKDIN
EISEIQAGYLETYLGDQADTDIDFVLTNNAKVRWLNVNKHDSAWCGLILDKWEE
SNKRFKGRDILDTYVLSDDYPGFK"
11506..12039
/function="unknown"
/notes="ORF9; seems to have no homologs in other

```

```

closterooviruses"
/codon_start=1
/product="19.6 kDa protein"
/protein_id="AAC40713.1"
/db_xref="GI:3090881"
/translation="MKLLSERVILRLSKSLRTNDHLVLILKEALINYNASFTDEG
AVLRDSRESIENFLVARCQNSQSRVWKALITNTVCKMSIETARSLGLLILVADSSV
SLDEAKSKIDKNFRLRRKRGKYYSGDSDVAKVKYIILSGENRGLGCVDSLKLVCGV
RQGGNV/LGHLILISLG"
12046..12585
/function="unknown"
/notes="ORF10; seems to have no homologs in other
closterooviruses"
/codon_start=1
/product="19.7 kDa protein"
/protein_id="AAC40714.1"
/db_xref="GI:3090882"
/translation="MDLSFIIVQILSASYNNDVYALTYLNLINAYNSVDDTRWRAINDP
QAEVNVVKAIVATTATELHRTLIDSIDSAFYDQVGLVGIARGLLRHSEDVLEVI
KSMVLFVCGRGRGSKRYLGYLSDQCTNRYMMLTQAGLAAVEGADILRTNHLVSGNKF
SPNFGIARMLLLTLCCGAL"
12588..12698
/function="unknown"
/notes="ORF11; seems to have no homologs in other
closterooviruses"
/codon_start=1

Query Match      31.4%; Score 2107; DB 79; Length 13154;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4608 ggatgtgtacaacgaagcgactatcgccaccaggtctcacaacgcatctaccctttgtcga 4667
Db 1 GGATGTGTACACGAAGCAGCTATCGCCACAGGTTCTCAAACGCAATTACCTTTGTGCGA 60
QY 4668 taacttgaaaggaggagtgcggtctcttttctcaaaactgggtgagggtgtatacctataa 4727
Db 61 TAACTTGAAGAGGAGGAGTGCGGTCTTTTCTCAAAGCTGGGTGAGGGGTATACCTATAA 120
QY 4728 tgggtgtagcattgtttcatcagggtgctcgctcgcttagagagatattcttaacggcgaat 4787
Db 121 TGGTGTAGCAGTGTTCATCAGGGTGGCTCGTGCCTCCTAGAGGATATCTTAACGGCAAT 180
QY 4788 taagtaccacaagcgctcttcgaccactgttttagtgcagaagtacaaagatgggtgagcgct 4847
Db 181 TAAGTACCCCAAGCGCTCTTCGACACACTGTTTAGTCAGAACTTACAAGATGGGTGAGCGCT 240
QY 4848 accattccacgctgatgcagaggagtgctatccatcagataaacccctatcttgacggtcaa 4907
Db 241 ACCATTCCACGCTGATGACGAGGAGTGCTATCCATCAGATAACCCCTATCTTGACGGTCAA 300
QY 4908 tctcgtggggaagcaaacctctcgcactaaagtgcaggaagggtggttaaggtcatggtcat 4967
Db 301 TCTCGTGGGAAGCAAACTTCGACTAAGTCAGCAAGAGGGTGGTAAAGGTATGTTGTCAT 360
QY 4968 aaacgtagcttcgggtgactattttcttatgccttcggtttttcaaaaggagcgaacttgca 5027
Db 361 AAACGTAGCTTCGGTGACTATTTTCTTATGCCTTTCGGTGTTCAAAGGACGCACTTGCA 420
QY 5028 ttcagtaaacctccatgcagcaagggcgcatcagtttgacgttcagggtcaactcggcgcgt 5087
Db 421 TTCAGTAACCTCCATGCACGAAGGGCGCATCAGTTTTCACCTTCAGGGCAACTCGGCGCGT 480
QY 5088 ctttggtgtaggcagatgttgagttgagcggcggtgctcgatgagagtcaccagg 5147
Db 481 CTTTGGTGTAGGAGGATGTTCAGTTAGCCGGCGGCTGTCGGATGAGAGTCACACAGG 540
QY 5148 tgttccaaacagcaaccacagagccaagtgtaccagacaatcacacaaaatcgagg 5207
Db 541 TGTTCAAAACAGCAACACAGAGCCCAAGGTGCTTACCAGAAACAATCACACAAAATCGGG 600
QY 5208 gggcaaggctctatctgaggaagtggttagggaagtcgaaggaggtcgacatactcgat 5267

```

```
Db 601 GGCCAGGCTCTATCTAGAGGAAGTGGTAGGGAAGTCAAGGGGAGGTGCACATACCTCGAT 660
QY 5268 atggtcggaacaagattacgttaggaagtgtagtggtcagggtcgataataatccagtgat 5327
Db 661 ATGGTCGGACAAAGATACGTTAGGAAGTGTAGTGGCTCAGGGCTGATAATCCAGTGAT 720
QY 5328 ggctctlgaacctgactacaccccataagacatttgaagtgggttaaacacgggacacctga 5387
Db 721 GGCTCTTGAACTTGACTACACCCCAATGACATTTGAAGTGGTTAAACCCGGGACCTCTGA 780
QY 5388 agatccggtcggtgagctacttgaagtattgctgctataggcattgagagacatacaggcg 5447
Db 781 AGATGCGGTGCGTGGAGTACTTGAAGTATCTGGCTATAGGCATTGAGAGACATACAGGGC 840
QY 5448 gttgcttatggctagaaaattgctgcctactaccgcgaagtggttctgaaagtacctaa 5507
Db 841 GTTGTCTTATGGCTAGAAATATTGCCGTCTACTACCGCGGAAGGTGTTCTGAAAGTACCTAA 900
QY 5508 tcaagtttatgaatcactaccggggtttcacgltttacaagtcgggcacagatctcatitt 5567
Db 901 TCAAGTTTATGAATCACTACCGGGCTTTCACGTTTACAAGTCGGGCACAGATCTCATTTT 960
QY 5568 tcaattcaacaagaagcgttgctggtgtagagacctaccgttactcatagctgaaaaa 5627
Db 961 TCATTCAACACAAAGACGGCTTGGCTGTGAGAGACCTTACCCTACGTACTCATAGCTGAAAA 1020
QY 5628 aggtatcttaccagaagggcaaatgacgcggtgtagcttggcgcaaatctggt 5687
Db 1021 AGGTATCTTTACCAAGGGCAAGATGTCACCGGTGGTAGCTTTGGCGCACAAATCTGTT 1080
QY 5688 cgtatgcagcatatactgggtttccacgatgccattaatattgatagggtgcactgaaagt 5747
Db 1081 CGTATCGCAGCATATACTGGTTTCCACGATGCCATTAAATTTGATAGTGCACCTGAAAGT 1140
QY 5748 cgtcgtatcgcatagtggtggggaatcgtttaagtccttcgaataaagtctataatgc 5807
Db 1141 CGCTCATCGCATGGTGGGGGAATCGTTTAAAGTCTTCGAATATAAGTGTGATAATGC 1200
QY 5808 tccccaggtggcggtaagacgacgcttagtggaacgaattcgttaagtcacccaatag 5867
Db 1201 TCCCCCAGGTGGCGTTAAGCAGCAGCTTAGTGGACGAATTCGTTAAGTCCACCAATAG 1260
QY 5868 cacagccaccattacggttaagtgggaagttctgaggacataaaatatggcgggtgaagaa 5927
Db 1261 CACAGCCACCAATTACGCTTAATGTGGAAAGTCTTAGAGCATAAAATATGGCGTGAAGAA 1320
QY 5928 gagagatccgaatttggaaggtctcaacagtctaccacagttaactcagggtggtaaa 5987
Db 1321 GAGAGATCCGAATTTGGAAAGGTCTCAACAGTGTCTACACAGTAACTCCAGGGTGGTAAA 1380
QY 5988 cttatcgtcaggggaatgtataaaaggttttgggtgaggtgcacatgatgcata 6047
Db 1381 CTTTATCGTCAGGGGAATGTATAAAGGCTTTTGGTGGATGAGGTGCACATGATCATCA 1440
QY 6048 aggttactacaactaggcgtcttcgcaaccggcgctcggagggcctctttttggaga 6107
Db 1441 AGGCTTTACTACAACACTAGGGGTCTTCGCAACCGCGGCTCGGAAGGCTCTTTTGGAGA 1500
QY 6108 cataaatcagatcaccattcaaaacaggaggaaggttttagatgagattgctgctttt 6167
Db 1501 CATAAATCAGATACCAATTATAAACAGGGAAGGTTTGTAGATGGATTGCTGTTTTT 1560
QY 6168 tgttccaaagaaggaaagcgttgtatatacacttctaaatcgtacaggtgccgttagatgt 6227
Db 1561 TGTTCCAAAGAGGAAGCGTTGTATACACTTCTTAAATCGTACAGGTGCCGTAGATGT 1620
QY 6228 ttgctacttgtttctcfaatgaccogtaaggggaaacggaaaagtgttaccctgaaaaagt 6287
Db 1621 TTGCTACTTTGTCTCTCAATACCGCTAAGGGGAACGGAAAAAGTGTACCCCTGAAAAAGT 1680
QY 6288 cgttagcgttaaggacaaacacagtagtaagatcgctgtccaaaggccaatggaaacac 6347
Db 1681 CGTTAGCGGTAAGGACAAACCAAGTAGTAAGATCGCTGTCCAAAAGGCCAATTTGGAAACCAC 1740
```

```
QY 6348 tgatgacgtagctgaaataaacgctgacgtgacttgcttgatgacccagttggagaagtc 6407
Db 1741 TGATGACGTAGCTGAATAAACCGCTACCGTGTACTTGTGCATGACCCAGTTGGAGAGTTC 1800
QY 6408 guatagaagaggtcgtgaaagggaagaaagaaacacacacagtcgatgacagtgcatga 6467
Db 1801 GGATATGAAGAGGTCTGTGAAGGGAAGAAAGAAACACCCAGTGATGACAGTGCATGA 1860
QY 6468 agcacagggaagaaacatttcagtgatgtgattgttttaggacgaagaaagccgagtaactc 6527
Db 1861 AGCACAGGGAAGAAACATTCAAGTGTGTGATTTGTTTAGGACGAAGAAAGCCGATGACTC 1920
QY 6528 cctattcactaaacacaccgcataacttctggtttgctgcagacacacacgcgtcactggt 6587
Db 1921 CCTATTCACTAAACACCGCATATACCTTGTGGTTTGTGAGACACACACAGCTCAGTGGT 1980
QY 6588 ttatgcgctctgagctcaaatgagctgacataaagtcgacacatatatttagcagacgcgtc 6647
Db 1981 TTATGCGGCTCTGAGCTCAAGTTGACGATAAGGTTCGCGCACATATATTAGCAGCGGTC 2040
QY 6648 acctcaatcagtatccgacgcttcttcacacgcttcgcccggctggtgttcttcgaagg 6707
Db 2041 ACCTCAATCAGTATCCGACGCTTGTCTCACAGTTCCGCCCGGCTGTTGCTTTTCGAGG 2100
QY 6708 tatatga 6714
Db 2101 TATATGA 2107

RESULT 3
AF283103 LOCUS AF283103 14861 bp DNA VRL 17-SEP-2000
DEFINITION Pineapple mealybug wilt associated virus-2 polyprotein (ORF1a) and RNA-dependent RNA polymerase genes, partial cds; and hydrophobic protein p5, heat shock protein 70, p46, coat protein, diverged coat protein, p20, p22, and p6 genes, complete cds.
ACCESSION AF283103
VERSION AF283103.1 GI:10179918
KEYWORDS Pineapple mealybug wilt associated virus-2.
SOURCE Pineapple mealybug wilt associated virus-2
ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae; Closterovirus.
REFERENCE 1 (bases 1 to 14861)
AUTHORS Melzer M.J., Karasev,A.V., Sether,D.M. and Hu,J.S.
TITLE Nucleotide sequence, genome organization, and phylogenetic analysis of Pineapple mealybug wilt associated virus-2
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 14861)
AUTHORS Melzer,M.J., Karasev,A.V., Sether,D.M. and Hu,J.S.
TITLE Direct Submission
JOURNAL Submitted (28-JUN-2000) Plant Pathology, University of Hawaii, 3190 Maile Way, Honolulu, HI 96822, USA
FEATURES
    source
        1. .14861
        /organism="Pineapple mealybug wilt associated virus-2"
        /db_xref="taxon:136234"
    gene
        <1..5520
        /gene="ORF1a"
    CDS
        <1..5520
        /gene="ORF1a"
        /codon_start=1
        /product="polyprotein"
        /protein_id="AAG13938.1"
        /db_xref="GI:10179919"
```

NVNAHVCNPMWPKDGGRRVRESWYLSDDGYKVCNVESLYKEVAAAKVTSCGREDPC
 GVSADAGFVDVYDQVSHQADMEKKNMRTFSLALMYPVELLAADGEIYLETDLVLY
 KREGDCRVSVGVTGESYTHSYRCISAYETAPVAVARGTVEKVQYEGYRGYHHISL
 CLSDSSASVQRIHRTSFHGSIKVIMPLRNLGVIISFERMYDIDFVDRIYAYSJNT
 IANLENRTPEYAGNVRAQKTHVITGVVHSHRVDITSNDLWGLVDFVAVOAKDKRRK
 SLDTYSLVKAMGDFLSVTINWVRLHRYSTTLFRLGMLWSSILAFNEDFRSLVAVPEN
 YLYRVKSVCLNLSRVLPLVTLNFCDAAEETQRLQDRIKSRKYKVVSKVLYLDAISL
 MKKRCSDGEDRDITEEDVVSFIMNMHNLHGFGLRGARIPITVGLKAVLSNAS
 SSTTRVALKVLASAVSRPVATVGSAAETSAVTCYKCDWNIKOYSDHACSLFGD
 GRRCATMGDCWDAQSWTVDKHTSWSEWFLALDKARQFAIGTFKSLPSIVESEFR
 DGVRCFDSITGSGQITNSHLCATGAIGLTTAMCYKRISEIRAYASTVTVTNPHL
 SNAARLLSWDGVTLVCLSDAFGLWLVTRYNCLAQILKTHIEGRCFRYGAVALSSY
 IVMRGCYQAVYICAGPOAKVAMALATLGLIKGVVPTVVRNVVNGVSLGLLYGT
 LPETSPGVIIVPVNTCNVYACEATFDSRSLSTRNDESDDDASSSGVALDSVK
 SUDDSTIRIMTRKESOPLOGISTAPTEPAVHTGPEIIVELSPDLDKGAIE
 RYNTSSIDEVIELEPLVLVKGAVECVNMHPISGAVDINOASCTAPPAEFTPTK
 PRTNALSDDSVVPSKSMSTTLTSMRDSRSFRSNKWLGRDVPVIALDEQSEPVIGDVG
 AFGEAPVVEYILYVATGLFRLYSLVNAARLIRVNDOKRVVFPVAVIHESCPGISEVS
 PTHIIPSKSQGDEVEDLSRVVFSVNRKRFETVAALKMADAATERYLVCDENVPHNA
 LNLVAAILHCAAGKVGKFSKTVKVCVNAPEGCGGTITKLIQYFDRGSGSCIMTANRG
 SAINDITIESIDAANASKAASNNVSGVESIDNYVCARTVNSQIMNCKGVMMNTCALV
 DMYLMLHGLLMLGVFSSGARRAIFYGDIQIPFINREKCFYSKEGVCPGKDEIYT
 SESYRCPADVCWMSLSLKAQAGSNRYLKGVCNOREVLRSLSKRPVVAEVOIOLEA
 DAYITFKOCEKVKVRLRAKRAVGRDKVFTSHEAQGMTEGRRVVLCLRSATDSDSESEP
 HILVALSRHTQSCVYATLSSKLADKVGAAIDSVTRKEYSDTVLKTVEVASALPRAD"
 <5468..7147
 /note="64.8 kDa; expressed via a +1 ribosomal frameshift"
 /codon_start=1
 /product="RNA-dependent RNA polymerase"
 /protein_id="AAG13939.1"
 /db_xref="GI:10179920"
 /translation="VIRYLPLRRRYFELIERCSWOODLERVLARCSNHFYVYVNS
 FLEEMVPSVDLYRFFEDDFSDHEEFLINSCLRDNSVNKLTRYENIYFISRSNI
 GPKRNTKLNFENRNFDDCYGDCDFADALVEKLVNFRFLGNRLFEIQSD
 VVCANAAASNNIDSPSTSGYKALLSALGYEYTPDGMRSRYKLVKSDAKPKLDETPL
 MKYVQNTIVYHDIRAITSFSOCFVOMVERLYVTDKVLVHGMDPSNLAKRIRADI
 GDINKYCYELDISYKDSOGALMKDVEQRLRLGLHEEIIDMEFCGEYDCLVSMWT
 REFETISGAQRSGGANTWLGNTIVYMLLSLLEESHYDVIYVSGDDSLIFSTEPLO
 DLHTLYTNGFDCKLLNNTPYFCSKELVQCKDLCYFVPDPFKLVFKHICKSTSVS
 DUHERFMSFDVTDLVSEDVVAAECLVWKRYHNTNTYAAICVYIHLRANFRQFLR
 MYLCTPALISICNGMNSFVSFKLIAKHLNLFELGNKYKDVVPIFDKTRAE"
 7772..7915
 /note="5.3 kDa"
 /codon_start=1
 /product="hydrophobic protein p5"
 /protein_id="AAG13940.1"
 /db_xref="GI:10179921"
 /translation="NLDAFTAITIIASLILAFLELLILFIVLVVYNYSRMHSSMRSY
 GAA"
 7929..9554
 /note="59.1 kDa; HSP70"
 /codon_start=1
 /product="heat shock protein 70"
 /protein_id="AAG13941.1"
 /db_xref="GI:10179922"
 /translation="MEVGIDFTTSTICYSAEGASGVLSFGSPYIETQVPIRADGT
 GYSIVKPKALYNAPVGRLYNPKRWGVNAYELDSYVLKLPVHRVEYFDKGSVML
 GIGIGEPRTSVTDIISLFKALIKEAEQSTGLRVTGAVTVVPADYNFKRSFTINC
 MKDLGTPVRAVINEPTPAALYSLSILQEKDLFLSAFDGCGTFDYFVKRLGDVYCVL
 LSGDNFLIGARDIDRAVAEAKVARGESIDTATLSLFAASIKEEVTNRPRAKTHVVL
 VDGKLIIFTSEDINDIVRPEARALHIYEQAQRYPHETSVAVLITGSSALQCVQEA
 LPAISKYDSKVPFDGRASDSYSAKIYCDILAGASKLRLDTLNTLSDELNRPV
 IVFSGKVPISERTITFTNGRKTMYGYEEVRSYLNALTFRGEYISNVEGNRTDS
 AFYVSSVSDGLSVSYNGTLLKNDLVPSPPTVFSKNLEYLSNIEKVANEGIPEYARQFM
 ALYGQRISREETLADVGAFAKEHKIVENYSKRWL"
 9792..11003
 /note="46.4 kDa"
 /codon_start=1
 /product="p46"
 /protein_id="AAG13942.1"
 /db_xref="GI:10179923"
 /translation="MHRRSALTQILEYSNNLLTVQSLFRKLIYEIDPMVLSSESKE
 ALQATANMTAVDKATSAVSLSKATSVSEIKGDTYTPVVVVKHMERVGAEEKLFP

AVIKAFLVDFSEVTNFMPTDTSLNIKSDVYSAYAQDTEEYLQBSIAKRIDTQFFKKWN
 VRFKQSLDMTYENRWAVYSEVDDIKYLDKFFNYSFNPVSVPVYVRIESSNP
 HELKNYFSNFRHGARSASSCRAPSIMLEWVLLIDSNTVETRLAPLAKVITLLWNSP
 IYGTNRULTRKKRYRFINLRNPKGRVDMQAVDDVANKRSVSSGVNIARYVCRYSSV
 TYSRKVLGITNNRNSLMDIEGILLAYDTVDALPVALVPKDWLRSYARACEYIRFNSN
 VTRGGEYGSNT"
 11079..11987
 /note="33.8 kDa; Cp"
 /codon_start=1
 /product="coat protein"
 /protein_id="AAG13943.1"
 /db_xref="GI:10179924"
 /translation="MAQNTVAVVEGTILESLTAPKPRFRVATSDVGKYVDSKYSRVST
 GVATERDRLPAIETELLATIPTASTDKGVIPEVKRSSNKPETDVIDDVTLLNPR
 KNVNLISGKTVKPVNPGPLISREIAIRIGEALKECHKQVMGSDSSDLDLATYFHL
 IOLATITFSKNSKEYKEDYIETQKKIYIKDSEVVERAAMNSGYENPFQYMRFY
 TSSSTLTUNGKITPNERMTAHGVPKQFFAYTDFIDPDYSILMHSAINAYNLTRIQ
 AFKNKIASVNTMHTNYQLNQAQVSG"
 12019..13494
 /note="55.8 kDa; Cpd"
 /codon_start=1
 /product="diverged coat protein"
 /protein_id="AAG13944.1"
 /db_xref="GI:10179925"
 /translation="MEFRIPAVEGSTFRLSKLDDKRYKIDLNGSLKWRVPHGER
 YDAYVRAITVDVDDPDMTKSDTIELDITVPILTVDYTTSGYIIHFEFIEDQCLRVG
 LGHNTYGLGRHGSYKSDGNASSETMLDSNFVSNWLSQYKILFSLVGRIVVKID
 NYVFRTPAVPOINPKILRIVTRLSTNSNLKYGTVPQAFKARFGLKFGANVTWVDS
 KYVRLVPDPDRAFPNPROYVDMVPDESUDKDVPTPIPSQVNTNATPSAAIQELNK
 VLSADKILIESRCKAILEFLPNATQNEADIDERYLDEQLSKVNTALSALCVELMGD
 KSRGALETLLIAMIQLCVTVSYVKNMIKKEYVETTYTRKLIYYSYLAIRSCIDKAV
 GSGFGEALNRQYMTYFTYTVYVAMRAGLVTPNYSAAKHGVPKRFYSYLAIRSCIDKAV
 YSQYDELAKSLAHAYAIKLAIKRGSDESSEVNTYSLGNNV"
 13484..13999
 /note="19.7 kDa"
 /codon_start=1
 /product="p20"
 /protein_id="AAG13945.1"
 /db_xref="GI:10179926"
 /translation="MEFRPIEVYEPENGKLSFNEINYEETTLTAETAYRLCGYSTGF
 TEGDEVTTNSLWLNAPAREPREYVDLIGVPTFYNNLRELLDIIPLNRTIKSVTLRR
 LPSDSGEVRIVLKNVIEKGGLPVSVEILPEVKGYNILKVYSWERDNRGIVKKSLL
 DATILLPKIPY"
 13996..14580
 /note="22.3 kDa"
 /codon_start=1
 /product="p22"
 /protein_id="AAG13946.1"
 /db_xref="GI:10179927"
 /translation="MSEELKKSADGMSVYHCLTLIALGEKITTEGRVELLINRLWFT
 HLSDDGKRMHYDVVENILTFQAQHRITIPQHTSVFLKYNVGNLINVDGYTSLILALE
 EFLARSDLELREQAVSEFGDGGFVPSOVVELYAKHNSKISETGVRRLEKKPLRDK
 DVFRRPKPEPSERDILLSAFVCIITDELYTRNCRK"
 14577..14729
 /note="5.6 kDa"
 /codon_start=1
 /product="p6"
 /protein_id="AAG13947.1"
 /db_xref="GI:10179928"
 /translation="MNTNAKKYLLSLILLIFVVAIISGETRSLLLGLILFLVGLIQEII
 KOINEY"
 BASE COUNT 4144 a 2718 c 3763 g 4236 t
 ORIGIN

Query Match 5.3%; Score 358.6; DB 79; Length 14861;
 Best Local Similarity 50.3%; Pred. No. 2.le-92;
 Matches 883; Conservative 0; Mismatches 874; Indels 0; Gaps 0;

QY 1153 gtccgtggcgatagacacggttggaaggagatcttggcctccgctgcgaatgccctt 1212

DB 757 GTGGCGGGTGAGAGGCCCGCGGTACAGACTGGTACGATCTATCGGCAAGCGCTC 816

QY	1213	aatcaagaggaggtctatttcggtcgttcgagtagtaccacatagactgggtattaag	1272
Db	817	AACCAGGAAGAGCTCAGCGTGGAGTTGTCTCATCTCACTCAACTCAAAAGTTAATGCTGAAA	876
QY	1273	gaccaatcggcattgcttttcccaatttggacacagaaatttgagatattgttttctcaaaag	1332
Db	877	GACCAAGGCCCTGTGTAAATCACTATGATACCCAGCTCTACAAAGATTTTAAAAATCGT	936
QY	1333	gacgcaatgatctcgcaaaaaaacctcacataggtgcgatgtgttctgaagcgcgaggaa	1392
Db	937	GATGAGGTACTGGGAAAGAGAGAGTGTCTACATTAAGTCTCAGCTTGCCACAAAGAAGA	996
QY	1393	agggaagaagctgaggaactctttccagagcttttcgatacagttctccgactcgttcag	1452
Db	997	AAGGACCTTTTACGTGAGAGTTTCCCGAGAGTTAAATTTGGTATTCTTGATTCGGTGCA	1056
QY	1453	agttagtcacccatctcgctaataatgcatacgcagagcgttcttcaatggaactctttccagagg	1512
Db	1057	AATGAGCACCCCAATGTGCAATAGCGTACGCTAFTGTGTTTAAACGTGCTATTGTGCATCAGCG	1116
QY	1513	tgTggTaaTgTgTcttcttcgataattggggagagcttcacgttatcatgtcaaaagctgc	1572
Db	1117	TATAAAGATGTCAAAATTTGTGAATGTGCGTGGCAGCGAGAGTTTATTCACGTGCTATTGTGCATCAGCG	1176
QY	1573	catgtgaactgtcatgtatgcaatccagctcctagacgtttaagatgtgaagcggagaatc	1632
Db	1177	AACGTTTAACGCTCATGTTGTAAACCCATGTTGATTGGAAGGACGCGGACGACGGGTA	1236
QY	1633	aatgagatcctcttcttccacagctggggagagattcgtaagtgctccagtgacacttcta	1692
Db	1237	CGAGAGTCTATGTACCTCTCAGACGACGGCTACAAGGTGTGAACGTGGAGTGTGTTGTAC	1296
QY	1693	actgaagcggcttcaaaagctgtgcttactgtagtcgagaatcgcagaaactgcgattct	1752
Db	1297	AAGGAGGTTGCCCTGCTCAAGTAACATCTCGCGGCGGAGTTTCAGAGATGTGCTGTG	1356
QY	1753	agagccgatgcgggttttatgtgtgagtgatgtacgatatatccccgcagcaggtagcagag	1812
Db	1357	AGCGCTGATGCAGGGTTTTTGTGTGATATACGATATGCTCCGTGCAGCAGTTGGCCGAT	1416
QY	1813	gctatgataaagaagggtgcgctggttttcgacataagctcttatgttccccgtggaagtgc	1872
Db	1417	GGAAATGGAGAAGAAGAAATATGCGCACTTTAGTCTCGGCTTTAATGTACCCCGTAGAGCTG	1476
QY	1873	ttgacggtTaaaggtgaagtttacttctggagaactcgatcagttggtgagagggaaggt	1932
Db	1477	TTGGCAGCTCAGCGTGAATTTACCTTGAACCCCTAGAGCTGTAGTGAAGAGAGAGGCG	1536
QY	1933	gattacctggcctacaatgttggtcagtggtgagatgatgaacattcctctctaac	1992
Db	1537	GATGTGTGCGCTATAGCTGGGCGTTACAGCGAGAGCTACACTCATAGTTATCGTTGT	1596
QY	1993	gtaagcgggttttccactttcttatgtacgcacttcgtccgggaacgtgtttaagcta	2052
Db	1597	ATATCGCTTATTTTACCCGCTCGTGTGTGGTGTAGGGGTGGTACTGTGTTCAAAGTG	1656
QY	2053	gagatagagggaTaccggttggttaccatcatctcactatgtgtgaggtcctcagaagtca	2112
Db	1657	CAATATGAGGTTACAGATGGGGTACCATCATATATCTTTGTGCTTAGCCGACAGCTCT	1716
QY	2113	cctggaaactgaggttacgtataggtcgttgggtcccgctgttcgtgggcaaatcgctggtc	2172
Db	1717	AGTGTGTAGGCTTAGTGTGAAAAGGACATCCCGTACCCTTCACGGAAGAGCATAGTT	1776
QY	2173	tccataccgttgtagctggttctagtggtctctttaagacaatagctcctcatcggac	2232
Db	1777	ATGATTTCCAGATTTGAATTAATGCGGTGATCTTCGACGCGTATGGTACGATACTGAT	1836
QY	2233	tttgtcagcaggaTctattctcactacgcgtccaacactataggacattcagagatagaacq	2292
Db	1837	TTTCGTGATAGGATTTATCGGTACAGTCTAAACACGATCCCAATCTCGAATCTGACG	1896
QY	2293	tttgagatgcgcgttggggcgttcaggtgcgcaaaagaccatgtcattacagggagtcgc	2352
Db	1897	TTTCAGTACGCCCTAGGCAACGTTTCGGGCACAAAAGACGATGTCTACTGGAAACCGT	1956
QY	2353	gttgtccacagaaggttatgatatttctcctgtagtatgtatggtggttttagttgcgtgtt	2412
Db	1957	GTGGTGCATAGTAGATAGATATAACGAGTAATAGCATTTGTGGGGTTTTGTGTGGGGTT	2016
QY	2413	atggctcagcgcgattaaagatagaggcgaagatttcgcctcctataacattataaagcc	2472
Db	2017	AGCACATCAAGCCGTGAAGATAGGCCAAGAGCCTTGATACGTATTCTTTGGTAAAGGCT	2076
QY	2473	agtgaaggagtgctgcgcgggggtcttcaagctcttcttcagaccgttaggcgattgtttt	2532
Db	2077	TGGGATGGAAACTTTTGTGTCAGTAATAATGTATGGTGAGACACTGTTGAGGTACAGT	2136
QY	2533	tcgaacgcagtcctccgctctatgctaaggccaatgggtgcacgataaactcaacgttttggag	2592
Db	2137	ACCACGCTCTTTAGAGGTCTTATGTGGAGTTCCTCGCGCTTAAATTTTGATACTTTTTCAGG	2196
QY	2593	acgttatgtctatgccagagcgtttcatccgtataagtagcgttgcgtctgttctgtttacc	2652
Db	2197	TCCTCGTGGCAGTGCCTGMAAATTACATATATAGGGTACCGAAGAGCGTGTGTTGAAT	2256
QY	2653	atttgcaactctggagcttcagacaggtttggagctcaggggtgccttctgatatttcgaag	2712
Db	2257	TTACGAAGCGTATTACCGTTAGTGACGCTCAATTTTGACTGTGCAGCTCGGGAACACAGT	2316
QY	2713	gagacctcggtaggaactgaagaatagtcgcttgcgctctctctcagggcctatcgctg	2772
Db	2317	CAACGACTACTTCAGATATCCGTAGCAAAAGGTATTAAGTACGTTTTCGAAGGTTCTCGTT	2376
QY	2773	gaagattcaatgaaggtcatgaagccaatgaagacaagaagatggaacccctcgcaatt	2832
Db	2377	AAAGATGCAATTAAGCCTAATGAAGAAAAGCGGTGTTCGATGGTGAGATCCGATATA	2436
QY	2833	actgaagattctgtatatcggttcaataatggggaagcttcttaacgtccactgtacgag	2892
Db	2437	ACTGAAGAGGACGCTCGTGTCAATTTATTATGAGTAACATGCACAATTTGCATCTTGGGTTT	2496
QY	2893	gcaggtctctctggcgg	2909
Db	2497	GGTGGTTACGAGGTGG	2513
RESULT	4		
LOCUS	AF195822	12394 bp	RNA VRL 09-MAR-2000
DEFINITION	Grapevine leafroll-associated virus 1 helicase (HEL) and RNA-dependent RNA polymerase (POL) genes, partial cds; and p7, HSP70-like protein, p55, coat protein (CP), p55 (CPd1), p50 (CPd2), p22, and p24 genes, complete cds.		
ACCESSION	AF195822		
VERSION	AF195822.1	GI:6653489	
KEYWORDS	grapevine leafroll-associated virus 1.		
SOURCE	grapevine leafroll-associated virus 1		
ORGANISM	Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae; Closterovirus.		
REFERENCE	1 (bases 1 to 12394)		
AUTHORS	Fazeli,C.F. and Rezaian,M.A.		
TITLE	Nucleotide sequence and organization of ten open reading frames in the genome of grapevine leafroll-associated virus 1 and identification of three subgenomic RNAs		
JOURNAL	J. Gen. Virol. 81 Pt 3, 605-615 (2000)		
MEDLINE	20141373		
REFERENCE	2 (bases 1 to 12394)		
AUTHORS	Fazeli,C.F. and Rezaian,M.A.		
TITLE	Direct Submission		
JOURNAL	Submitted (18-OCT-1999) Plant Industry, CSIRO, Hartley Grove, Urrbrae, Adelaide, SA 5064, Australia		
FEATURES	Location/Qualifiers		
source	1..12394		
	/organism="grapevine leafroll-associated virus 1"		

gene	/db_xref="taxon:47985"	gene	6890..7858
CDS	<1..1199	CDS	/gene="Cp"
	/gene="HEL"		6890..7858
	<1..1199		/gene="Cp"
	/gene="HEL"		/codon_start=1
	/product="helicase"		/product="coat protein"
	/codon_start=3		/protein_id="AAF22742.1"
	/db_xref="GI:6653490"		/db_xref="GI:6653495"
	/translation="LRYVDANFKVVAAPEDGTTIPDLEYVFLVSTGTFVPRRNQAV		/translation="MASVISQNDNDYNNVGGNIVVPRQISLGSFGCTSAITYVTPTGAI
	LEQDAVLCRACDELLEFHDAMNLCGRVRLAKRAMVGYMKDVRISAVNSPPGGNTTRL		AVYLKTHYKPEAGTAANYPVVGLPDESVEFKGPGYQSRLSRSTQRFQDSEV
	VBYQFQKVKRACIAAAGTADINAAIRARECKKEKPELVAKTANSVNSHPRPNSH		RRMGDKLKRNTKNIATIFNPMNIIFFPKDMEVSVLPTGCLVTPAVATLSTEL
	VGLIDVYVHLHGMFOLTVYSGVKEVYIAGDKNQIPFINREKTFVTPNEAVEAEQ		KNLCAVMGNTQKSLTDFPLMQLMLTFSTPDTESEKEEYFNLYCNGERKLTYEK
	IDYDLSRCPADVCYVLSMTDARGKMYPNGFENGDRPLRSEKVPYIATPEDAL		VGAVVKGAGSGSTFENPMQYARLFSATAVHLLNCKLKPNEKYAMQHGVPKRFLPYT
	LEADVYLWTONKAEKOKAVAKMEVWAGKRPDVITTHAQQGTIENVILRUKA		FFCRPSYQFSNDALRAWQLAAESAFGRKSNVTSSVLRNTSELK"
	DPIFSRKHPIVVALSRHRTSMKYAVLSSKMTDTISKLIDGTSAQKVSVDVLLQQLRN		7881..9383
	DRFRD"		/gene="Cpd1"
gene	<1147..2730	gene	7881..9383
CDS	/gene="POL"		/note="coat protein duplicate 1"
	<1147..2730		/codon_start=1
	/gene="POL"		/product="p5"
	/codon_start=1		/protein_id="AAF22743.1"
	/product="RNA dependent RNA polymerase"		/db_xref="GI:6653496"
	/protein_id="AAF22738.1"		/translation="MATLVAPIVGDKEVLAALLRNLPNDESLDNLNAGRNDDFS
	/db_xref="GI:6653491"		RVEAYRLNLSIGYTNVSVSISIKPIKETMGLRVVIEPRDGILLIHTIRFDRKNL
	/translation="VFCCSCSGCTIAPESIECFSVAPOLEKFIKAKVTSHRVNL		IEQVYQRKGVSVLKSRPDASIVNFHSERNEKMTSMNKQTRLMETQVOTGSSLGT
	FLECRVPGITSUNTNWEDDEFETSGLETIVETATISDNFPKYKITDRPRSQYVRSAI		SOIKVPESLITVLVLESKVKVSCDTLSGMIPTEVSSIIRFDGTIDYKQIDRERKIF
	RPKRNSLCNLTVPESRNFNADRGDVSDFVAAAEALASLFTNWVDGSKLASCVGD		TKQNAVAASWSDRSMAKQSLWVLRFSVYDLAAMTRYQLMVKADAKPLDSTPL
	QOYVGTONTVYHDRAITAMFSLHFTQAVERLVKLVHSMVAYHGMSTDDFSEVCERL		PVKPTIDEVATLLSATENTDLPIGRSKLNVSLFDVLELTGYNPSMVNDVDVNVNT
	RDISGVYVELDISKYDKSGQACMDVERLLILIGLVAESVDVAFCEGYDSVYWGK		KILVEYSLYKSESSENALAGIMIOGALTWTSANLKGDKERKKVDVSLNGKNTYDF
	NELVSVGAQRSGGANTWGLTVLMTLLAISLKGFEPLVVCVGGDDSLIFSKNELV		NELRQIIRSVPPSPKFNPIRQYRMWFTTITSLKSGVVVVPNYHMYARHGVTSQFIP
	INDRILESFGDLKLTCCQVPYFCSKYLVRTPDHCYFLPDPFKLFLKLSLEREPED		YGFQICILLPSYNNRRDDKVAALARAQAVALANKRAGKTLTYNFSSELGS"
	LLFVETSVDTLTRFGCDENAVQQLVENDAQRYGFNPYAAFLVHVLAANFVQFKR		9387..10709
	LFDFDAGAFGKKLSKATR"		/gene="Cpd2"
CDS	3525..3704		9387..10709
	/note="ORF2: hydrophobic"		/gene="Cpd2"
	/codon_start=1		/note="coat protein duplicate 2"
	/product="p7"		/codon_start=1
	/product="p7"		/product="p50"
	/protein_id="AAF22739.1"		/protein_id="AAF22744.1"
	/db_xref="GI:6653492"		/db_xref="GI:6653497"
	/translation="MDLRQFSHELLTVSLFVLVLCFAVYFVRAIRHCCAKIEDKP		/translation="MEIVEAVDSRGNAFRIYDVKYALLAVYSRSMLLCSRVRFSSEFR
	VGPTNSPRYVAQP"		SLYQWIGRGKNTLELGVDFGGSHGAKLPFPYVYGGYIYNEQMTPEISSDLDIVD
CDS	3710..5341		VELSLTKSRKVVTAASVNGIDVPCIGYTLSENGSLRFGHEIIFHVNDEELRYLVKST
	/codon_start=1		RNRRIEGLYEAIVKVDGQVLLGDVEMADTRINRAYSNKKGVPSPVHTNVRLKTL
	/product="Hsp70-like protein"		FOITASVKTIDVRHGKLEKAMLYYEVFDKNLEDAVKVCSYSEIPNTVAGLTNVDFKG
	/protein_id="AAF22740.1"		IDNITLSVSGECAYLDKDVAKITLALLKPLIVDNKSVYIDKRLRLILIIOCAVYVG
	/db_xref="GI:6653493"		TRADFVKGSTEVISYKSGSVGLDFISVRHVHIIACGGINVTNPVROYRWRWSTVTI
	/translation="MEVGLDFTGTTSCFSIPTQDDSGCVSLVNSPFVPTQIFGSD		QLIKGIVPNPNIVAARRLTNNNTWLSFDYILLDARYNDRKEKILML"
	MYSGHRAYSFVAGKPGGLYNRWRWGVDSFNHVRKRLNLPNIEYKINNGEISI		10789..11358
	GVGNTNAPLMRVVDVLFVKGILLETEAVGKAVTGVCVTPAEYVSKFRSFLGA		/note="ORF8"
	LEGLKPLRLNEPTSAALYGAVGGSIRETYAVDFCGGTLDISFISRFNNVSVL		/codon_start=1
	FKGDNFLGRDTRDAIVNFKRKRKIGNIDAGILSMIADLKERKICTNGGTQTOVQ		/product="p22"
	KTSNGLTSMVDNLNAVSEPPIDRAVKYFAEGAEDLKRCPVCLVTGTSVALPLVR		/protein_id="AAF22745.1"
	PLENLVYSSVAYDSQTRLSVAIGAKIYGDLTGQSLRLDITVTSQSLBELSGFT		/db_xref="GI:6653498"
	ELVIFPKGAPVSEYTSQVSGSTMEYIGFEGESNRKTWNEIAFGTDRPSNERKN		/translation="MEFAPVCYIITKNDGGCKKLLSRDLNIDFNSDDYQBELLE
	DKVYBISVDGKLLKSDVGRKLTPLPAPVLSAHSYRYVSSMKKYLQLENNYVDM		KKPFVKSYNYLLEWAEKPCYLSNAMSLKMDLAHMTKASIESDSIDGKLGKVLTAVE
	FSELGDKTISIDDVYNDTGAYFDKSLVNFSLSK"		VQSPKTKTJPFDSVNGVLLAVGYEGLNMEKLRARSIALVMQDGSYVVLGCEVHTS
CDS	5338..6786		SPYFIECLRCRLASTMSGDVFVDALEFI"
	/note="ORF4"		11402..12034
	/codon_start=1		/note="ORF9"
	/product="p5"		/codon_start=1
	/protein_id="AAF22741.1"		/product="p24"
	/db_xref="GI:6653494"		/protein_id="AAF22746.1"
	/translation="MNSLVWFTGTTDYSVLKPLVDFDKAEGRRVVNYIEVIDRGNK		/db_xref="GI:6653499"
	PYLIRGVSPDSNFELIAWAAVAGPENSIIYAFETILSMRLNWKTYANOGSLKDMPI		/translation="MASLIPKYVGSFSELCAHIVEISNADSVHLLYEALKLNRLWGL
	TKKKHFLVTEQRAARALPNATNEAQAQLFLNLGVLPNRPNEEDVSGGYIPLY		SFSTFVEETQKELEQTPFGQULRFAREHTVAEIDTFMRMYGGEYUREANTMKDLTA
	QVSSWATRIEGRSPLLAKFVRYITIIIDAYSVTRQKIEVCYSYIPPELLALLF		MLHVLVDYTKLEASTYGLSRDRKRGFSRTLKRVTAKTOELQTSIDKTSGLDGAEVK
	RTNDNTPAYKIFSLYLASELQSDPGKDYSETRVIAFTRALARELLEVEYNIYSVI		LFAIALSLAEEDSHGSPKDVVARTNRLSSYLLKDKSFTTFQFCAIASNLV"
	SEREIIRNFRIQDFEHLQSLMLGLVADTAAKTIGISNVDSVLRRLVGLCURAWS		
	ENOHISINWTGALYLILYFISVGTNYLRKKTIDEMTKINGVVRKVGFTSLKSI		
	AEASVDGRNVPREISSVFAGITLQRLQSGIDTVVWPDVVLNPNCLGFTDLHCGVIN		
	LAPAFYKIDIEVKRIRNSAHWVGKYGKRP"		
		BASE COUNT	3542 a 2219 c 3190 g 3443 t
		ORIGIN	

Query Match	3.3%;	Score 220.2;	DB 79;	Length 12394;
Best Local Similarity	50.3%;	Pred. No. 3.8e-52;		
Matches 580;	Conservative	0;	Mismatches 558;	Indels 15; Gaps 1;
Qy	5568	tcattcaacaagagcgcttcggtgagagacctcaccgtactacatagctgaataa	5627	
Db	38	TAATGCACCGAGACCGGAACACTACTATTTTCGACCTGGGAATACGTTCTTCGTATCTAC	97	
Qy	5628	aggtatctttaccaagggcaaaagatgacgcgggtggtagcttttgggcgacaactctgtt	5687	
Db	98	AGGCACCTTTGTGCCACGGCGGAATTTGAACGGGTCTTATCACGACAAAGATCGCGTACT	157	
Qy	5688	cgtatgcacgatatactggttttccacgatgccattaatattgatagtgcaactgaagt	5747	
Db	158	GGTGTGGGATCAACTGCTCGTTTTTTCATGATCAATGAATTTTCGCGGGATCGTTAGGTT	217	
Qy	5748	cgctcgatcgcgatggtggcgaaatcggtttaagtccttcgaataataagtgctataatgc	5807	
Db	218	GGCGAAAGAGCAATGCTCGGAATAATATGAAGAGATGAAGAATTTCTGCGGTGAATC	277	
Qy	5808	tcctccaggtggcgtaagacgacgcttagtggaacaaattcgttaagtcacccaatag	5867	
Db	278	ACCGCTTGGTGAGGTAACACGACGAGCTGGTGTATGAATACCTTTGGAAGGAACGCG	337	
Qy	5868	cacagccaccattlacggtcaatggtgggaagttcttgaggacataaaatggtgggaaga	5927	
Db	338	AGCTAAAATCCGACGACCAACACCGGAAGTGTAGCCGACATCAATGCGGCTATTTCGAGC	397	
Qy	5928	gagagatccgaatttgaaggtctcaacagtgctaccacagtttaactcagggttggttaa	5987	
Db	398	AAGAGAGGGAAGAAAGACGAGCATTTAGTACAAAGACCGGAATTCATGGGTCATATA	457	
Qy	5988	ctttatcgtcaggggaatgtataaaagggttttggtagaggtgcacatgatgcacca	6047	
Db	458	CTCCATCCACGTCGCACTCGCACGTTGGACTTATCGATGAGGTTTATATGCTACATA	517	
Qy	6048	aggettactacaactaggogtcttcgaaccgcgctcggaagcgctcttttttggaga	6107	
Db	518	GGGAATGTTTCAGCTTACTGTGCTCTCTATGGAGTGAAGAGGTGATAGCGTACGGGA	577	
Qy	6108	cataaatcagataccattcataaacaggaggaaggtgttagagtgagttgctgtgtttt	6167	
Db	578	CAGAACCAATTCGGTTCATCAACAGAGAGAACATTCGTGACTCCAAATGAAGCTGT	637	
Qy	6168	tgttccaaagaagaaggggtgttacacttctaaactacaggtgctcggcttagatgt	6227	
Db	638	GGAGTTCCGTCAGGACAGATAGATTATACTGACATTTCTTACCAGATGTCACGACAGT	697	
Qy	6228	tlgclacttgttgcctcaatgacogtaaggggaaagcgaaagagtttaccctgaaaagt	6287	
Db	698	TTGTTATCTTTTCTTCGTGACTGATGCACGTCGAAGAAATAATGACCCAAACGGAGT	757	
Qy	6288	cgtagcggtaagagacaaccagtagtaagatcgctgttccaaaaggccaattggaaccac	6347	
Db	758	GTTCGCCGAATGGAGACGTGAGACCTCTACGCTCGTTCGGAAGAGGTGCGCATCGCAACGCC	817	
Qy	6348	tgatgcagtagtgaaataaaacgctgacgtgtacttgtgcatgacccagttggagaagt	6407	
Db	818	GGAGACGCTTTACTCTATGAACGCGACCTGTATCTGACGATGACTCAGACGAGGAGCC	877	
Qy	6408	ggatatgaagaggtcgttgaagggaagga-----aaagaaacacacagt	6452	
Db	878	TGAAATGCACAAGGCTGTGCAAAAATGCAACTGTGTGCTGGGAAGAAACACCGGATGT	937	
Qy	6453	gatgacagtgcataagagcagggaaaaacatactcagtgatgttggtattgttagaagaa	6512	
Db	938	GATTACTACTCATGAGCGCGCAAGGTGAAGACATACGAAACCGTGCTTGGTGAGACTGAA	997	
Qy	6513	gaagccgatgactcccttactactaaacaacgcgcatatacttgttgttgcgagaca	6572	
Db	998	GACGCGGATGATCCAATTTTTCGCCGAAACACACATATGTGGTGGCGGCTGTCCAGGCA	1057	
Qy	6573	cacagcctcactggtttatgccgctctgagctcgaagttgacagcataaggttcgagcata	6632	

```
EGVTSKGGPREAMLLGDGVGVHFRVDRKADRLAQLAQSVEKVLVISPRLVDRDDWAKISKHRY
ELEPKEVLTAFCAGSGASHMCQRTILKHCSEKVLVISPRLVDRDDWAKISKHRY
VTFEAFMDYDCKDIVIDEIGLLPGYIDLVIHAHQPTLVLGDPLOSTYHSKRDN
VLEASQEDVNRVRKLPYLCYSHRPLRNCKLFEIECMGAESKRVYRSNRLKDEP
TTCATRAMKEEGSGWYTVSETGSLFKSLIYLDEHMAKKEDEDMVALTRSGEIG
IHWTPALKKKLLTNAKSTLLKKVLAGETVRSSEIVAMVRKHIPETTVLPFEESRLAETV
DYEARLADGPYLLSLALYDEIMEDIETEEPTVLEPTKTHALSTKMMELAPFDLKA
KEHRQHTFAGTGTQIENGSGEYGDPMTHKALYLHPTSDDTAFMMSVKRLFRN
YEANRRKYTCGHIGHOMPSVEKDYQLKEISLPELERCENEFMKRIEKSTGLIEK
HAGRSDDPMSNLYKIFLQOQTKMEKRGVDKAGQTIACFAHSLVCRFGPILROYE
KALRELPKLMYISOKRYMDLKWATWESMGTDSDYEAFDRSQDEKVLDELEV
LRFLLWPELIDREYBELKLMGMCALDLAVMRFSGEFGTFFFNTVCNMVFSWCRVHID
RNTPMCFAGDDMYSPGILLRKVDYATLIDLTLKAKVHISEEPLFCGWRMSPFGLIKE
PNLLIDRWKIALRSNLSLCLVAYAEASFYRLSEHLYDVNLDVDAQOELVREIVIK
KILLPKKISDLFSEDECERHSQDEDFLSNDVARLYRIE"
5179. .5712
/genes="ORF2"
5179. .5712
/genes="ORF2"
/codon_start=1
/product="unknown"
/protein_id="CAA53183.1"
/db_xref="GI:1405616"
/db_xref="SPTREMBL:O67705"
/translation="MTSRDCTGLSFLHGHGSDSGSLGSLSVOCVNLISDLKS
LQYSIDSILYILGGEARERFIYRIFRPHGIGEALQLGVKKSCLNSPRSLIALID
DLISLRGSAFLPSDLGAVGELVATFHSRSLSDVLYNNKKVYTRSVQAEQDYSYV
ARRFGYKGLILRATRY"
5654. .6490
/genes="MP"
/notes="ORF3"
5654. .6490
/genes="MP"
/codon_start=1
/product="movement protein"
/protein_id="CAA53184.1"
/db_xref="GI:1405617"
/db_xref="SPTREMBL:O67706"
/translation="MSQEGSLGTKASSPEPQDIKVFHVKRSTRDLTLNKSLSHRGDVY
NTELEKPEPRKTCVTHKVDKGRVDCDLDDINDEEFLYHVGCIV
ALMPHGKLVGSVSVLDRVLDGASRISRTLMDSKPLSADKPPGYFISTDLIN
GYTLHLSITTTDLOQVDGVHPFSVOLMSIGRFGCEMDKTRYAITETSKMLHONILNTE
GGELIIPRGVQVQKVPDITLMPVEVETIKKFKLKTNGTLRQESGRDKDNRNRYGVGES
TN"
6414. .7010
/genes="CP"
/notes="ORF4"
6414. .7010
/genes="CP"
/codon_start=1
/evidence=experimental
/product="coat protein"
/protein_id="CAA53185.1"
/db_xref="GI:488390"
/db_xref="SPTREMBL:O67707"
/translation="WAHYAKRVEIRAIIEELVLAQAQPTDDASESGYDRNMYLNTLFG
YIALGTSKDLIHYGEVDIVGPASKKGTIDPRGMVSVSELVGRMTLSVAVSEGPVK
GATLRQMCPEPQAYDELVMAENGTYSQLATKMTGRSGKEPQVNFDFASGLDLKAL
TLOEATVIOAMHSRLRFTREGAGVFNQAQSSIGEQAIVEI"
7015. .7281
/genes="ORF5"
7015. .7281
/genes="ORF5"
/codon_start=1
/product="putative RNA binding protein"
/protein_id="CAA53186.1"
/db_xref="GI:488391"
/db_xref="SPTREMBL:O67708"
/translation="WTHFSRVGRPLLVSVGARMNVCKGAIMHNKDKCKSSISSHK
LDLRFVKEGRVALTGETPPVRYTWKMYETEYHIVETSDDED"
2195 a 1590 g 2014 g 1550 t
```

```
Query Match 1.3%; Score 87.4; DB 79; Length 7349;
Best Local Similarity 50.9%; Pred. No. 1.6e-13;
Matches 208; Conservative 0; Mismatches 201; Indels 0; Gaps 0;

QY 4665 cgataactgaagggaggagtcgctgttttctctaaagctgggtgaggggtataccta 4724
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1883 CGAGAACTGAAGGGCGCGAGGTGGCTTCTATTAGCAGGCACAGCAAGGAGTACAATA 1942

QY 4725 taatggtggtagcattgttcatcaggggtgcctctgccttagagagatattcttaacgc 4784
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1943 CAATGGGGCTCCCATCGTTCCCTCGGTGGATGAGGCTCTAAATGAGCTCACACAGGA 2002

QY 4785 aattaagtaaccaagtccttcaccactcttagtgcagaagtaacaagatgggtgaggg 4844
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2003 ACTGGGCCTAGATGATAGCTACGATCATTTGCTGATCCAAAGATATACACGCAGGGGCGAG 2062

QY 4845 cgtaccattccagctgacgagagtgctctatccatcagataaacctctcttgaagg 4904
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2063 TATAGGATCCACGCGATGATGAACCATGCTACCTACCGGGGTTCCCTCGTGACCGT 2122

QY 4905 caatctcgtgggaagcaaacctctcgaagtgagggaaggggtgtaaggtcatcgtg 4964
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2123 CAATTTACACGGCGACGCTACTTTTGAGGTGAAGAGAAATCAATCCGGGAGATCGAGAA 2182

QY 4965 cataaacgtagcttcggtgactattttcttgccttcggttcgttttcaaaaggacgact 5024
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2183 GAAGGAGCTACACGACGGGGATGTGTACGTGATGGGTCCAGGAATGCAACAACTCAAA 2242

QY 5025 gcattcagtaaacctcctgacgaagggcgccatcagttgacgttcagg 5073
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2243 GCACAGATTACCTTCACACGGGATGGCCGCTGCAGCATCACCCCTTCGG 2291

RESULT 6
AITASPV 3306 bp RNA VRL 05-FEB-1999
LOCUS Apple stem pitting virus (ASPV) RNAs for putative viral
DEFINITION RNA-dependent RNA polymerase, triple gene block ORFs and coat
protein, complete cds.
D21829
D21829.1 GI:487651
coat protein; triple gene block; viral RNA-dependent RNA
polymerase.
SOURCE Apple stem pitting virus (ASPV) (individual isolate PA66, strain
ASPV) cDNA to genomic RNA, clone No.1-278.
ORGANISM apple stem pitting virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Carlavirus.
REFERENCE 1 (bases 1 to 9306)
Jelkmann,W.
AUTHORS Direct Submission
TITLE Submitted (28-OCT-1993) to the DDBJ/EMBL/GenBank databases. Wilhelm
Jelkmann, Institute for Plant Protection in Fruit Crops, Federal
Biological Research Centre; Schwabenheimer Straase 101, Postfach
1264, Dossenheim D-69221, Germany
(E-mail:ba9degenius.emblnet.dkfrz-heidelberg.de, Tel:06221-85238,
Fax:06221-861222)
REFERENCE 2 (bases 1 to 9306)
Jelkmann,W.
AUTHORS Submitted (28-Oct-1993) to DDBJ by:
TITLE Wilhelm Jelkmann
JOURNAL Federal Biological Research Centre
MEDLINE Institute for Plant Protection in Fruit Crops
COMMENT Schwabenheimer Str. 101
Postfach 1264
D-69221 Dossenheim
Germany
Phone: 6221-85238
```


Fax: 6221-861222
E-mail: bat9ed@genius.embnet.dkfz-heidelberg.de.
Location/Qualifiers

FEATURES
source

1..9306
/organism="apple stem pitting virus"
/isolate="PA66"
/strain="ASPV"
/db_xref="taxon:35350"

CDS

34..6585
/note="putative"
/codon_start=1
/product="viral RNA-dependent RNA polymerase"
/protein_id="BAA04853.1"
/db_xref="GI:487652"

/translation="NALLSRRTAAEEVIASTSEOSRISTQAVLALTIVKDKHDLFN
YALPELAKMRFLNSGIYSPISYRSHSPVCKTLENNILFNILPSYLDNSFYLSIKK
NKVDLPKRMHEDQWETINIRYISSIDKTRYGGFFHVSPSKISAKFKCDRTGFEDDA
SLDILPCMCRRGACRRFFHDEHLHWTKREALITFDHWKPEVMLASIFVPEIPILGAK
ESLNPCWCTFTRVGDLVFFPDGESEAVIQPVAGSYLRITGKITTPSGDILFQDLKK
SFSFHLISITKGEAIGOKMRFFNGEFAVAMGLNPLRKVESCLPISKNTILKITYR
LRTLKKPDQSAKMLSOVCKDPNGYEIKFEPEFSKCLCKDCTLNTNIPDMKRLVOG
FELFLPNPISNFKVQQLHLNDPIETLEEFENSINTESLSLWNKDDLEFVNLFIFG
TDNFVDFAEAWGTKDVNITIVHSPSYLVSKFESYDHFHSLSVLSKLSALPIRGA
KIVLSLYPCVVEAFSESRVTNLANVNIANLRACFATDMLRIFEGILLKECKRAQ
GKMRKRFHFLGIRWFLVDVSNQWFLPPCRDGLTARSFDDQFIKGCQRONSLSHNGR
MSLRQVLKPKLQALFDVSELSIIHNWEMENAPEAGSTLDAGIKPTSGPLEVPIENA
RCNLAPCKDLNCFIQPADVNSLHGNLVLFDFIGSGKRGASFYSRDLKGYSYTGFSH
VSRGWPALDFLSDNKIPLNFIYNOCLVOEYSTGHLSMHRKDDESIYDINHQVLTVNY
SDQALFICELSGSEIFLSPQMULMFPGFQKEHRHGKISPSKGRISLTFLKLTREGD
SGQVIEQVVTICDHGSDDRALKALERSHOSGGRPAVELEGHEREKVNSDSDSAP
VOEFIQIDSSLLYALKLSGLSNVYVCMCLCNPLWKNELRFSLEARLDLAFQA
GLIQLIDLCVLKACAVNRILISELPTHPVLRGTMHIVDLDDRESIRGDVKEGFSFG
PRRWKMSCTDILMALFLPKMTLGGELRSHEDCELSDLTEKHLGSCSVILSRKFEF
DLFHSPDVEADGNCWHWSVGLLIGVDGEYLRKLIHQAKKGKVCPLRSKOLEGNTWA
EREAVYFCSHGIRLNYITREECTWIFKPEVILKAAATLILICQDNHFRAPCPVNGCVI
RAISSANREVDVLAJLGKPAHEDLFEVAGRGFSIFDLTRLEIFESIGSVDTGG
ELIMNENGRIPAEFLSKEHLAHITLSRRKFSPIVSDLNARVNSAMRFLAINGAEV
DYRPSIDRASTLDSFEIGATGVLCOGIEKAQDLAKLILPELVHERKLIIMLTGFGC
GKSLFKFIEKSPCKALTEVSPRLAESINHDGLARVGSKTKGSKDLKNVRVKT
FELFILHDSIEKGVTVIDELQFPFGYIDLIIILGPNVNIILAGDPCQSDYDCSS
DRHIFAGSEDMRLISGRSYKFNILSRFRNPFVYGRPLCNKRLTRLDEEEYTLW
DSIQFMSMKRDCVLSVRSFEEKIIVAAHIGIKMKCTITYGESTGLNFPQKGLIVTY
ESALTSRDMWTLSRFSDIHFINGMVTWDAITHEVGKPLHRFETTKRACNDIID
LPLGRPELIEGQSOQADGEGVREAKLVGDPLWTKIFLGNPNPDEIETADEVEAAD
WEKTHIPMSLEAVRAOVKHILSREDREFRGDITTEQFTDDHSHKNGOELTNAER
YEAIYPRUKGTDATFLMAVKRLSFSPPAAEHAHLRAKPPGKFLLOTFLKRVPLNS
SHDEKMQEAVHAFEEKLSKSMATIEHNSRSCEDVPDKALIFMKSOQLCTKFDNRF
RSKAGQTLACQFHSVLCRFAPYMRYSIESKVTPEVLPNKLYIHSNGKNIIDDLAAVYTSK
FNGVCTESDYEFADSQDHFILAFLEYMKFLGLPSDLIADYTFIKTHLGSKLGSFAI
MKRTGEASTFLFNTMANMLFTFLRYDLNGREAICFAGDOMCANSLKVTNRFNSFLDK
IKLKAQVOFTAPPTFCGWGLCEHGVFKKPKDLVLERLOIARTRNLNCIDNYAIEVSC
AVKMGENLNYLTPOEVDAHNCVRFIVQHNHLLKSNIRDLFKGESLPSA"

CDS

6685..7356
/note="triple gene block as found in potex- and
carlaviruses"
/codon_start=1
/product="ORF2"
/protein_id="BAA04854.1"
/db_xref="GI:487653"

/translation="METVLSLLNEFGFERTVEPLSDPIVVHVPVSGSKTTLIKOALIR
NNNIEAVTGVPEKAIHIGTYIKKARQOGRGNYSILDEYLSGEYSYFNCFLSDPY
QNHGDLRAHFTGRGSHRFGROTQVILRLDGLNYIASSEKEDIVEKKNIIFQLIPEGVII
CLEKGVEDFLKHWISVEYKFPCCQVRGATDITVFIHEKPLEELVGLDFVALTRHSKL
VLVSN"

CDS

7358..7720
/note="triple gene block as found in potex- and
carlaviruses"
/codon_start=1
/product="ORF3"
/protein_id="BAA04855.1"
/db_xref="GI:487654"

/translation="MPFAQPPDYSKVSVPFIAGVIAVAVLFTLTRSTLPVQGDNIHNL
PHGGNYQDGTKRISYCGPRDPSPPSSLSISGTPMIITGIIIFLIFAIYVSEKWSRGSR

CDS

RCSCCVPGAPACTATVHE"
7629..7841
/note="triple gene block as found in potex- and
carlaviruses"
/codon_start=1
/product="ORF4"
/protein_id="BAA04856.1"
/db_xref="GI:487655"

CDS

/translation="MFPRSGLGLAFAAAVAVLVLLLLAQQLYMSNSSOCTIVITGESV
SVVGCYSEAFIELVKGLKPIYHPLG"
7930..9174
/codon_start=1
/evidence=experimental
/product="coat protein"
/protein_id="BAA04857.1"
/db_xref="GI:487656"

BASE COUNT 2567 a 1864 c 2175 g 2700 t
ORIGIN

Query Match 1.2%; Score 80.8; DB 79; Length 9306;
Best Local Similarity 52.5%; Pred. No. 1.3e-11;
Matches 228; Conservative 0; Mismatches 197; Indels 9; Gaps 2;
QY 4675 aaaggagagagtcggtctcttttctcaaaagtcgggtgaggggtatataatggtggt 4734
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2188 AAGGTCGTGGGGGTAGTTTACTCGAGGGATCTAAAGGGGTACAGTTACACTGGATTTC 2247
QY 4735 agccatgtttcatcagggtggcctcgtgccctagagagatatcttaacggcaataggtac 4794
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2248 AGCCATGTCTCGAGGGGTGGCCTTCTCGGATAAATTTCTTAGTCACACAAATTT 2307
QY 4795 cc---aagcgtcttcgaccactgttagtcgagaagatacaagatgggtgagggcgatacca 4851
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2308 CCGCTAAACTTCTACAAACAGCTGTTTGGTTTCAGGAGTATAGCACTGGTCACGCACTTTC 2367
QY 4852 ttccacgcctgatacagagagtgctatccatcagataaacctatcttgacgctcaatctc 4911
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2368 ATGCAATAAGATGATGAGAGCATTTATGACATCAACATCAAGTTTAAACAGTTAATTAT 2427
QY 4912 gtggggaaggcaaaccttctcgactaagtcgaggaagggtggtaaggtcactggtcataaac 4971
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2428 TCTGGTGATGCCATTTTCTGTATAGATGCCCTGGGGTCAGGTTTCGAATTTCCATTAAAGC 2487
QY 4972 gtactctgggtgactattttcttgccttcggttttcaaaaggacgacttgacttca 5031
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2488 GGACC-----TCAAAATGCTTCTCATGCTGCTTTTGGATTTCAAAAGGAGCATAGACATGGG 2541
QY 5032 gtaaacctccatcgacgaaggcgagcagctcagcttgcagtcaggcgcaactcggcgcgtcttt 5091
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2542 ATCAATACACCTTCTTAAAGAGTAGATCTCTTTGACACTTTTCAGGCTGACTAAGGAGGTGAT 2601
QY 5092 ggtgtaggcaggat 5105
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2602 AGCCAAGTCCCGAT 2615
RESULT 7
AF034622
LOCUS
DEFINITION
Tobacco rattle virus RNA polymerase, 29 kDa protein, and 16 kDa
protein genes, complete cds.
ACCESSION
AF034622
VERSION
AF034622.1 GI:28522336
KEYWORDS

SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Tobacco rattle virus.
Tobacco rattle virus
Viruses: SRNA positive-strand viruses, no DNA stage; Tobravirus.
1 (bases 1 to 6790)
Sudarshana,M.R. and Berger,P.H.
Nucleotide sequence of both genomic RNAs of a North American
isolate of tobacco rattle virus
Unpublished
2 (bases 1 to 6790)
Sudarshana,M.R. and Berger,P.H.
Direct Submission
Submitted (17-NOV-1997) Plant, Soil, and Entomological Sciences,
University of Idaho, Room 242, Agricultural Sciences, Moscow, ID
83844-2339, USA

FEATURES
source
1. .6790
/organism="Tobacco rattle virus"
/isolate="ORV"
/db_xref="taxon:12295"
/chromosome="RNA1"
202. .5325
/note="readthrough protein"
/codon_start=1
/transl_except=(pos:3763. .3765,aa:OTHER)
/product="RNA polymerase"
/protein_id="AAC02066.2"
/db_xref="GI:8050240"

CDS
/translation="MANGNFKLSQLINVDMSAQRSHFFDLMLTKPDCEIGOMQVRVYKDDMIREKTKDPVIVHEVLSQEKONLMEIYPEENIVFKDDKNVHGFAAAR
KIQALLLDRLVPALQVDDIGGQSWFWTRGEKRIHSCCPNLDIRDODREISROLFLT
AIGDQARSKROMSENLWYDQFRENIAAPNAVRCNNYQGCRCGFSGDKKKAQY
AIALSLYDFKLDLMATYVKKTKVYVHAAFLAPESMLVDEGLPVSVDGYKKNGK
IYFGFEKDSFYIHDWEVKYLLGPKVSYQGNVFEFPWQVGRDGMFLFSYIAGV
PRSSLSQSYRIIYISRMENWVVPIDLVSTRELKVKDLFVEKQFMDKLDIYAR
LSDQQLTISNKSYSLSNNVLFINGAAVKNOSVDSRDQLLQALLOLLVKEQVAPVM
RELRAILTETAPITSLTDVLGILSRKLWKQFANKIANGGVFGVWGTLGIFPKVPLT
WAKDPNGPELCYENSHKTVIVLSVYVYAGGITLMRRDIRDGLVKKLCLDMFDIKRG
AHVLDPNFCRYEINDFSSLSASEGETVLPDLSEVKARSKDLQOQKEADEFL
SAKFSNCGSVYTPSPVSGSRGLGLLDENVLTQARVSGKRVKDEIMEQFL
SLGIDTEADISAFSAECERGETSGTKLCKPLTPFGFENLPAVKPLVSKGKT
KRVDFQVNGERLPRKVPVSGDNSVDARREFLYLDAERVAQNDIEIMSLYDSRGV
ITGGONYPHGIVGDMVEMKNCIRPVVTEHAVVQPDKRMDDSCYLKVAWRCML
VNDFAVERSDYIVCDOTYLCNNRLI LONSLDLPANCSFELVDGPGCGKSTMI
VNSANPCVDVLTSGRAADDLIERFASKGFECKLRKRVKTVDSFLMHVCDGLSGDVI
LHFDALMAHAGVYCAQIAGKRCICQGDQNSIFKPRVSOVDLRFSLVSKGFDIV
TEKRETPSPADVAVLNKYTGDRTHNATNSMTKIVSKEQVSLKPGQYITFL
QSEKELVLLALRVAKVSTVHESQGETFKDVLVTRKPTDDSTARGREYLIIVALS
RHTOSLVYTVKEDDVSEIRESAALTKAALARFVETVLXRFERSRDFVRHHEGPC
AVPDSGTITDLEWYDALPQNSLRDSSLDGYLVATDCNLRDNLVTKSGNKKKFA
EKEARLKPVIETAMPDKTKTQLESLLALQKRNQAPDLOENVHATVLIETMKKLS
VYDVGKIRADPIVNAQWRWNOSTAIQAKVYADVRELHEIDYSSYMYKSDVK
PKTDLTPOEYSALQTVVYHEKLIINSFGPIFEKNERKLDAMQPHFVNTRMSSDL
NDRVFLATFEAYDVEIDMSKFSANRFLQLQLEIYRFLGLDEWAAFLWEVSHIQ
TTVDTQNGMAHVIYQOKSGADTYNANSRDTLCALLSELPLEKAVMVTYCGDSDTL
APRGTQFVDPCKPLATKNFECKLFKYDVPMECGKFLFKTSCEYEFVDPDKVLLKL
GKSKIDVOHLAEIYISLNSNRLNGYVWVSKLSVESVSDRYLYKGDSVHALCALWKH
IKSFALCTLFRDENDKELNPAKVDWKKQRAVSNFYDW"
5327. .6085
/note="29 kDa protein"
/codon_start=1
/protein_id="AAC02067.1"
/db_xref="GI:2852338"

CDS
/translation="MEDKSLVTLKKKTFEVSKFNSLGAIELEFVDRGRRKRPKYFHRRE
TVLNVHGKSKSEHLDQFQDYDKMKISYAFKIVGVLVVTSHLPADPGFIQIDL
DLSRLTEKKGKHTIQFRKARDNCVSAQYKVEYSISTOENVLDVWVKCISGVPVC
DCTYFPIBSVLWATDSTRILNVEELNSDY IEGDFTDQEVFGEFMSLKQVEMKTI
EAKYDGPVEPATPKSLSSSEDDVKNRKNKSS"
6110. .6535
/note="16 kDa protein"
/codon_start=1
/protein_id="AAC02068.1"
/db_xref="GI:2852339"

CDS
/translation="MTCVPKGCVNEVTVLGHETCSIGHANKLRKQVADVMGVTRCAE

Query Match 1.1%; Score 75; DB 79; Length 6790;
Best Local Similarity 54.1%; Pred. No. 6.3e-10;
Matches 153; Conservative 0; Mismatches 130; Indels 0; Gaps 0;
QY 6319 tcctgtccaaagcccaattggaaccactgatgacgtagctgaaataaaacgctgacgtg 6378
DB 3391 TCATGACGGTGAAGAGATTGCTCTAAAGACACAGGTTCTTTTGAACCCCTGGTCTCAG 3450
QY 6379 tacttgtcagatgacccagttggagaagtcgatatgaagaggtcgcttgaaaggggaaaaggga 6438
DB 3451 TACATAACTTTTCCTTCAGTCTGAGAGAGAGTTGTTGTAATTTTGGCATTGAGGAAA 3510
QY 6439 aaagaaacacagtgatgacgtgcatgaagcacagcagggagaaacattcagtgatggtgta 6498
DB 3511 GTGCAGCTAAAGTGTAGTACAGTACAGAGTCCGACAGGAGACATTCAAAGATGTAGTC 3570
QY 6499 ttgttaggacgaagaagccgcatgactccctattcactaaacacgcatactactgtt 6558
DB 3571 CTAGTCAGGACGAACCTACGGATGACTCAATCCTAGAGTCCGGAGTACTTAAATCGTG 3630
QY 6559 ggttgtcgagacacacacgctcactgctgtttatgcgcctctga 6601
DB 3631 GCCTGTGCGGTACACACACATCATTGTGTATGAACACTGTGA 3673

RESULT 8
AF166084 6791 bp RNA VRL 10-AUG-1999
Tobacco rattle virus RNA 134 kDa protein, 194 kDa protein, 29 kDa
protein, and 16 kDa protein genes, complete cds.
AF166084
VERSION AF166084.1 GI:5714670
SOURCE Tobacco rattle virus.
ORGANISM Viruses: SRNA positive-strand viruses, no DNA stage; Tobravirus.
REFERENCE 1 (bases 1 to 6791)
AUTHORS Visser,P.B. and Bol,J.F.
TITLE Complete nucleotide sequence of RNA 1 of tobacco rattle virus
JOURNAL Isolate Ppk20
REFERENCE 2 (bases 1 to 6791)
AUTHORS Visser,P.B. and Bol,J.F.
TITLE Direct Submission
JOURNAL Submitted (07-JUL-1999) Institute of Molecular Plant Sciences,
Gorlaeus Laboratories, Leiden University, Einsteinweg 55, Leiden
2333 CC, The Netherlands
FEATURES Location/Qualifiers
source 1. .6791
/organism="Tobacco rattle virus"
/isolate="Ppk20"
/db_xref="taxon:12295"
/chromosome="RNA1"
203. 3766
/note="putative replicase"
/codon_start=1
/product="134 kDa protein"
/protein_id="AAD48026.1"
/db_xref="GI:5714671"

```
RELREAILTETPTISLTDVLGLISRKMMKQFANKIAGVGFVGMVGTLLGFYPPKKVLT
WAKDTPNGPELCYENSHKTKIVFLSVVYVYIGGITLMRRDIRDGLVYKLCDFMDFIKRG
AHLVDVNPCTRYDINDFSSLYSASEGETVLPDLSEVKAKSDKLLQOKRKHIADEFL
SAKFSYSGSSVYTPSPSVVSGSSGSLGLLLEDNVLTQARVGSVRKVADEIRMEIQFL
SGLITEAEIRDVPAFAEACERGETSGTKVLCNLLTPPGFENVLPYKVPVLSYKGVTV
KRVDFQVWGGBELPKRPVPVSGDDSDVADAREFLYIDAEVAQNDIEMSLYRDSRGV
IRTGQNGPHGLGVDMVENKNCIRPVVTEHAYVSNPKRMDMDSGYLEVAVYERGMTI
VNDFAVERMSDVIICDQTYLNCNRLIIDLNSALDLPVNCSEFLVDPVGGCKSGMI
VNSANPDVVLSTGRATDGLIERFASKGFPCKLRRVKTVDLSFLMHCVGSLGTDIV
LHFDLMAHGAHWYFCAQIACAKRCICQGDQNOISFKPRVSOVDLRFSSLVGKGDIV
TKRRTYSPADVAALVANKYITGDVTHNATANSMTVRKIVSKEOVSLKPGAQYITFL
QSEKELYNLLARVAKVSTVSHESQGETFKDVLVVRKTPDDSDIARGRELYLVALS
REKLSLVTAKEDDVSEKIRESAALTKAALRAFVETVL"
join(203..3763,3765..5326)
/note="putative replicase"
/codon_start=1
/exception="readthrough protein"
/product="194 kDa protein"
/protein_id="A4D48027.1"
/db_xref="GI:5714672"
/translation="MANGNFKLSOLLNVDEMSAQRSHFFDLMLTKPCEIGOMMQRV
VVDKDDMTRERKTKDPIVHEVLSQKEQNKIMEIYPEFNIVFKDDKNVHGFAAER
KIQALLLDLRVPALQEDDDIGQWFSWYTRGEKRIHSCCPNLDIRDDREISQIFLT
AIGDQARSGKROMSENLMDYDFRKNIAAPNAVRCNNNTYHGCTCRGSDGKKGQAY
AIALHSLYDFKLDMATVEMKTKVHAAMLFAPEMLVDEGPLSDYGYMKKNGK
IYFGEKDPSESYIHDWEYKYLKGKPVSYOGNVYFEPQVGRDQTMFLFSYIAGV
PRRSLSSQYRYRIVISRWENMVVPIFDLVSTRELVKDLFVEKQFMKCLDYIAR
LSDQOLTISNYSKYLSSNNWVLFINGAAVKNQSVDSRDLQALLAOTLLVKEQVAPVM
RELREAILTETPTISLTDVLGLISRKMMKQFANKIAGVGFVGMVGTLLGFYPPKKVLT
WAKDTPNGPELCYENSHKTKIVFLSVVYVYIGGITLMRRDIRDGLVYKLCDFMDFIKRG
AHLVDVNPCTRYDINDFSSLYSASEGETVLPDLSEVKAKSDKLLQOKRKHIADEFL
SAKFSYSGSSVYTPSPSVVSGSSGSLGLLLEDNVLTQARVGSVRKVADEIRMEIQFL
SGLITEAEIRDVPAFAEACERGETSGTKVLCNLLTPPGFENVLPYKVPVLSYKGVTV
KRVDFQVWGGBELPKRPVPVSGDDSDVADAREFLYIDAEVAQNDIEMSLYRDSRGV
IRTGQNGPHGLGVDMVENKNCIRPVVTEHAYVSNPKRMDMDSGYLEVAVYERGMTI
VNDFAVERMSDVIICDQTYLNCNRLIIDLNSALDLPVNCSEFLVDPVGGCKSGMI
VNSANPDVVLSTGRATDGLIERFASKGFPCKLRRVKTVDLSFLMHCVGSLGTDIV
LHFDLMAHGAHWYFCAQIACAKRCICQGDQNOISFKPRVSOVDLRFSSLVGKGDIV
TKRRTYSPADVAALVANKYITGDVTHNATANSMTVRKIVSKEOVSLKPGAQYITFL
QSEKELYNLLARVAKVSTVSHESQGETFKDVLVVRKTPDDSDIARGRELYLVALS
REKLSLVTAKEDDVSEKIRESAALTKAALRAFVETVL"
/note="putative replicase"
/codon_start=1
/exception="readthrough protein"
/product="194 kDa protein"
/protein_id="A4D48027.1"
/db_xref="GI:5714672"
/translation="MANGNFKLSOLLNVDEMSAQRSHFFDLMLTKPCEIGOMMQRV
VVDKDDMTRERKTKDPIVHEVLSQKEQNKIMEIYPEFNIVFKDDKNVHGFAAER
KIQALLLDLRVPALQEDDDIGQWFSWYTRGEKRIHSCCPNLDIRDDREISQIFLT
AIGDQARSGKROMSENLMDYDFRKNIAAPNAVRCNNNTYHGCTCRGSDGKKGQAY
AIALHSLYDFKLDMATVEMKTKVHAAMLFAPEMLVDEGPLSDYGYMKKNGK
IYFGEKDPSESYIHDWEYKYLKGKPVSYOGNVYFEPQVGRDQTMFLFSYIAGV
PRRSLSSQYRYRIVISRWENMVVPIFDLVSTRELVKDLFVEKQFMKCLDYIAR
LSDQOLTISNYSKYLSSNNWVLFINGAAVKNQSVDSRDLQALLAOTLLVKEQVAPVM
RELREAILTETPTISLTDVLGLISRKMMKQFANKIAGVGFVGMVGTLLGFYPPKKVLT
WAKDTPNGPELCYENSHKTKIVFLSVVYVYIGGITLMRRDIRDGLVYKLCDFMDFIKRG
5328..6086
/function="involved in virus transport"
/note="putative"
/codon_start=1
/product="29 kDa protein"
/protein_id="A4D48028.1"
/db_xref="GI:5714673"
/translation="MEDKSLVT"KKKTFEVSKFSLGALTELPVDGRRKRPKYFRRRE
TVRLNHVGKKSHKIDRFQDQDYKMIKSYAPLKIYGVQLVYVTHSLPADTPGFIQDILL
DSRLTEKRRKTIQRFKARQDCNCSAQYKVEYSISTQENVDLWVKVCSIGEVPPKI
DGTYPFSIEVSLIWTASTRLNLEENSSDYIEGDTDQEVFGEFMSLQVEMKTI
EAKYDGYRPATTPRKSLSSDVKRANKNS"
6111..6536
/note="small cysteine-rich protein"
/codon_start=1
/product="16 kDa protein"
/protein_id="A4D48029.1"
/db_xref="GI:5714674"
/translation="MTCVLKGVNEVTVLGHETCSIGHANKLRKOVAMVGVTRCAE
NNGCVFCVYINDFYVNCGRSHLEKRRKRVETRNREIKWQIRRNQENMSATAK
KSHNSKTSKKFKEDREGTKPRFLRDDVDPFGIDRLFAF"
1986 a 1130 c 1743 g 1932 t
```

```
Query Match 1.1%; Score 75; DB 79; Length 6791;
Best Local Similarity 54.1%; Pred. No. 6.3e-10;
Matches 153; Conservative 0; Mismatches 130; Indels 0; Gaps 0;

QY 6319 tcgctgtccaaaggccaattggaaccactgacgacgtagctgaaataaaacgctgacgtg 6378
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3392 TCGATGACGGTGAGGAAGATTGTGCTAAAGAACACAGGTTTCTTTGAAGCCTGCTGCTCAG 3451
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 6379 tacttgtcatgacccagctggagagtcgatatagaagaggtcgcttggaaggagaaagga 6438
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3452 TACATAACTTTCTTCGACTGTGAGAGAGAGAGTGTGTAATTTGTGGCATTGAGGAAA 3511
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 6439 aaagaaacaccagtgatgacagtcgacgaagcacaggaagaaaacattcagtgatggta 6498
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3512 GTGGCAGCTAAAGTGAGTACAGTACAGGAGTCCGCAAGGAGAGACATTCAAGAGATGATGC 3571
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 6499 ttgtttagacgaagaaagccgactgactccctattcactaaaaaacgcgcatacttgtt 6558
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3572 CTAGTCAGGACGAAACCTACGGGATGACTCAATCGCTAGAGGTCGGGAGTACTTAACTCGTG 3631
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 6559 ggtttgtcagacacacacgctcactcgtgttatgcgcgtctga 6601
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3632 GCCTTTCGCGTCACACACAAATCACTTGTGTATGAAACTGTGA 3674
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 9
LOCUS MTRNAL 6791 bp RNA VRL 15-MAY-1998
DEFINITION Tobacco rattle virus RNA-1, complete genomic sequence.
ACCESSION D00155
VERSION D00155.1 GI:222133
KEYWORDS 194k protein; 134k protein; 29k protein; 16k protein.
SOURCE Tobacco rattle virus (strain:SYM) cDNA to genomic RNA, clone:24B,
25A, 31B, 4A, 17A, 25B, 13A, 543B, and MA61.
ORGANISM Tobacco rattle virus
VIRUSES: ssRNA positive-strand viruses, no DNA stage; Tobravirus.
REFERENCE 1 (bases 1 to 6791)
AUTHORS Hamilton,W.D., Boccara,M., Robinson,D.J. and Baulcombe,D.C.
TITLE The complete nucleotide sequence of tobacco rattle virus RNA-1
JOURNAL J. Gen. Virol. 68 (Pt 10), 2563-2575 (1987)
MEDLINE 88034943
COMMENT The RNA-1 of TRV contains four open reading frames coding for 134K,
194K, 29K, and 16K proteins. 194K proteins may be produced by
readthrough of the stop codon for 134K protein. Amino acid
comparisons of the 194K protein with the putative replicase of TMV
showed three regions of strong homology, suggesting that the 134K
and 194K proteins were similarly involved in virus replication.
FEATURES
Location/Qualifiers
1..6791
/organism="Tobacco rattle virus"
/strain="SYM"
/db_xref="taxon:12295"
/chromosome="RNA-1"
/clone="24B, 25A, 31B, 4A, 17A, 25B, 13A, 543B, and MA61"
203..5326
/note="position 3764..3766 is readthrough"
/codon_start=1
/transl_except="(pos:3764..3766,aa:OTHER)
/product="194k protein"
/protein_id="BAA00110.1"
/db_xref="GI:1707356"
/translation="MANGNFKLSOLLNVDEMSAQRSHFFDLMLTKPCEIGOMMQRV
VVDKDDMTRERKTKDPIVHEVLSQKEQNKIMEIYPEFNIVFKDDKNVHGFAAER
KIQALLLDLRVPALQEDDDIGQWFSWYTRGEKRIHSCCPNLDIRDDREISQIFLT
AIGDQARSGKROMSENLMDYDFRKNIAAPNAVRCNNNTYHGCTCRGSDGKKGQAY
AIALHSLYDFKLDMATVEMKTKVHAAMLFAPEMLVDEGPLSDYGYMKKNGK
IYFGEKDPSESYIHDWEYKYLKGKPVSYOGNVYFEPQVGRDQTMFLFSYIAGV
PRRSLSSQYRYRIVISRWENMVVPIFDLVSTRELVKDLFVEKQFMKCLDYIAR
LSDQOLTISNYSKYLSSNNWVLFINGAAVKNQSVDSRDLQALLAOTLLVKEQVAPVM
RELREAILTETPTISLTDVLGLISRKMMKQFANKIAGVGFVGMVGTLLGFYPPKKVLT
WAKDTPNGPELCYENSHKTKIVFLSVVYVYIGGITLMRRDIRDGLVYKLCDFMDFIKRG"
CDS
```

CDS

CDS

BASE COUNT
ORIGIN

AHVLDVENCPRYYEINDFFSLYSASESGETVLPDLSEVKAQSDKLLQOKKEADEFL
 SAKFSNYSSSVTSPPSVVSGSRSGLLGLEDNSVLTQARVGVSRKDDDEEMEQFL
 SGLIDTEAEIDEVVSFAFSAECERGETSGTKVLCKPLTPPGFENLPAVKPLYSKGTIV
 KRDYFQVGMGRLPKRPVVGSDNSVDAREFLYLDABERVAQNDREIMSLYRDYSRGV
 IRTGQYVPHGLGVGVENKWNWCIRPVVTEHAYVFQDPRMDWDSGLYEVAVWERMGL
 VNSDPAVERMSDVIVGCDOTYLCNNRLIDNLSDALDGPVNCSELPVLDVPGCGKSTMI
 VNSAPCDVVLSTGBAATDLDLTERFASKGFCPKLRVKTVDSFLMHCVDGSLGTDV
 LHFEALMAHAGWYFCAQIAGAKRCICOGDQNOISFKPRVSOVDLRFSSLVGKFDIV
 TDKTLPQEPADVAALVNYTGDVTHNATNSMTVRKIVSKQDSVSLKPGAYITFL
 NDRVKFLNTAAAYDFEYIDMSKDFKSANRFLHQLQLEIYRLFGLDEWAAFLVSVSTH
 TVTRDQNGMAHAYTQKSGOADTYNANSRDTLCALLSELPLEKAMVYTGDDSLI
 APPRGTFQVDPCKLATKNFECKIFKYDVPMPFCGKFLKTSCTSYEFVDPVKVLTNKL
 GKKSINKDVEHLAEIYLSLSDNSPALGNVWVSKLSSEVSVDRIYKGDSDVHALCALMKH
 IKSFTALCTLFRDENDKELNPAKVDWKKQAQVNSFYDW"

CDS

/codon_start=1
 /product="134k protein"
 /protein_id="BAA00111.1"
 /db_xref="GI:222134"

/translation="MANGNEFKLSQLLNVDMSAEQRSHFDFLMLTPDCIEIGOMQORV
 VVDKDMIRERKTKDPIVHEVLSQEQNKLMIEYFEFNI VFKDDKNMVHGFAAER
 KIQAALLDRPALQVDDIGQWSPWTRGEKRIHSCNPNLDIRDQREISQIFLT
 AIGDQARSQRKQSENELMWDQFRKNIAAPNACNNTYQGCTCRGFSOGKKGAOT
 AIALHSLDFPKLDMATVVEKTKVVAHMLFAPESMLVDEGPLSDVGYVMKKNGK
 IYGFGEKDFSFYIHDMEEYKYLKGPVSYOGNVEYFEPQVGRDTMLFYSRIAGV
 PRSLSSQYRIIYLSRNMENMVVIFDLVSTRELRKDLFEVQFMDKCLDIYAR
 LSDQUTISNYSLSNNWLPFINGAAVKNQSDVSDRLQLLAQTLVKEQVAPVM
 RELREALTETPITSLTDVLGILSRKLMKQFANKIAGVGFVGMVTLICFPYKVLIT
 WAKDTPNGELCYENSHKTKIVFLSVVAIGITILMRDRIDGLVKKLDMEIDRFL
 AHVLDVENCPRYYEINDFFSLYSASESGETVLPDLSEVKAQSDKLLQOKKEADEFL
 SAKFSNYSSSVTSPPSVVSGSRSGLLGLEDNSVLTQARVGVSRKDDDEEMEQFL
 SGLIDTEAEIDEVVSFAFSAECERGETSGTKVLCKPLTPPGFENLPAVKPLYSKGTIV
 KRDYFQVGMGRLPKRPVVGSDNSVDAREFLYLDABERVAQNDREIMSLYRDYSRGV
 IRTGQYVPHGLGVGVENKWNWCIRPVVTEHAYVFQDPRMDWDSGLYEVAVWERMGL
 VNSDPAVERMSDVIVGCDOTYLCNNRLIDNLSDALDGPVNCSELPVLDVPGCGKSTMI
 LHFEALMAHAGWYFCAQIAGAKRCICOGDQNOISFKPRVSOVDLRFSSLVGKFDIV
 TDKTLPQEPADVAALVNYTGDVTHNATNSMTVRKIVSKQDSVSLKPGAYITFL
 NDRVKFLNTAAAYDFEYIDMSKDFKSANRFLHQLQLEIYRLFGLDEWAAFLVSVSTH
 TVTRDQNGMAHAYTQKSGOADTYNANSRDTLCALLSELPLEKAMVYTGDDSLI
 APPRGTFQVDPCKLATKNFECKIFKYDVPMPFCGKFLKTSCTSYEFVDPVKVLTNKL
 GKKSINKDVEHLAEIYLSLSDNSPALGNVWVSKLSSEVSVDRIYKGDSDVHALCALMKH
 IKSFTALCTLFRDENDKELNPAKVDWKKQAQVNSFYDW"

CDS

/codon_start=1
 /product="29k protein"
 /protein_id="BAA00112.1"
 /db_xref="GI:222135"

/translation="MEDKSLVTLKKKTFEVSFNLGAIELEFVDGRRKRPKYPHRRRE
 TVLNHVGGKSKSEHLKLVDFQORYKMKISYAFILKIVGVQLVVTSLHPADTPGFTQIDLL
 DSRLETKRKGTIOERKACDNCVSAOYKVEYSISTOENLVDVWKGVCISGVPVC
 DCTYFESVSLWVATDSTRRLNVEELNSDYIEGDFTDQEVGEFMSLKQVEMKTI
 EAKIDGPYPATIRPKSLSSSEDDVKRASNKNSS"

CDS

/codon_start=1
 /product="16k protein"
 /protein_id="BAA00113.1"
 /db_xref="GI:222136"

/translation="MTCVLKGCVNEVTVLGHETCSIGHANKLRQKQVADVMVGVTRCAE
 NCGMPVCVVIINDFTFDVYNCGRSHLEKCRKRVETRNREIMQIRRNAENNSATAK
 KSHNSTKSKFKFEDREFGTPKRLRDDVPFGIDRLFAF"

BASE COUNT 1993 a 1126 c 1740 g 1932 t

Query Match 1.1%; Score 75; DB 80; Length 6791;
 Best Local Similarity 54.1%; Pred. No. 6.3e-10;
 Matches 153; Conservative 0; Mismatches 130; Indels 0; Gaps 0;

QY 6319 tcgctgtccaaaggccaattgaaccactgatgacgtagctgaaataaacacgctgacgtg 6378

Db 3392 TCATGACGGTGAAGAAGATTGCTCTAAAGAACAGGTTTCTTTGAAGCCCTGGTGCTCAG 3451
 QY 6379 tacttgtgcatagccaccgttgagaagtgcgatatgaagaggtcgcttgaaaggggaaagga 6438
 Db 3452 TACATAACTTTCCTTCAGTCTGAGAAGAGAGTGGTAAATTTGTTGGCATTGAGGAAA 3511
 QY 6439 aaagaaacacagtgatgacagtcgcatgaagcagcagaggaacacattcagtgatgtgta 6498
 Db 3512 GTGGCAGCTAAAGTGAGTACAGTACAGAGTCGCAAGGAGAGACATTCAAAGATGTAGTC 3571
 QY 6499 ttgtttaggacgaagaagccagatgactccctattactactaaacacgcataacttctgt 6558
 Db 3572 CTAGTCAGGAGCAACCTACCGGATGACTCAATCGCTAGAGTCGGGAGTACTTAATCGTG 3631
 QY 6559 ggtttctcgagacacacacgcctcactgctgttctatccgcctctga 6601
 Db 3632 GCGTGTGCGGTACACACACATCACTTGTGTATGAACACTGTGA 3674

RESULT 10

TOTRRNA1 6791 bp RNA VRL 07-JAN-1997
 LOCUS Tobacco Rattle virus (TRV) RNA-1 complete sequence.
 DEFINITION X061172 D00155 X03576
 ACCESSION X061172.1 GI:62132
 VERSION
 KEYWORDS replicase; subgenome; unidentified reading frame.
 SOURCE Pepper ringspot virus.
 ORGANISM Pepper ringspot virus.
 VIRUSES: sRNA positive-strand viruses, no DNA stage; Tobravirus.
 REFERENCE 1 (bases 1 to 6791)
 AUTHORS Hamilton,W.D., Boccara,M., Robinson,D.J. and Baulcombe,D.C.
 TITLE The complete nucleotide sequence of tobacco rattle virus RNA-1
 JOURNAL J. Gen. Virol. 68 (Pt 10), 2563-2575 (1987)
 MEDLINE 88034943
 REFERENCE 2 (bases 4484 to 6712)
 AUTHORS Boccara,M., Hamilton,W.D.O. and Baulcombe,D.C.
 TITLE The organisation and intervirial homologues of genes at the 3' end of tobacco rattle virus RNA1
 JOURNAL EMBO J. 5, 223-229 (1986)
 COMMENT *source: MAG1;
 [2] Data kindly reviewed (08-JUN-1986) by W. Hamilton [1] [2] merge data kindly reviewed (23-NOV-1988) by W. Hamilton.

FEATURES

source
 1. .6791
 /organism="Pepper ringspot virus"
 /strain="SYM"
 /db_xref="taxon:31750"
 /clone="34B, 25A, 31B, 4A, 17A, 25B, 13A, 543B,"
 203. .5326
 /note="put. 194 kD protein (AA 1 - 1707) put. replicase"
 203. .3766
 /note="put. 134 kD protein (AA 1 - 1187); put. replicase"
 /codon_start=1
 /protein_id="CAA29537.1"
 /db_xref="GI:62133"
 /db_xref="SWISS-PROT:P05080"
 /translation="MANGNEFKLSQLLNVDMSAEQRSHFDFLMLTPDCIEIGOMQORV
 VVDKDMIRERKTKDPIVHEVLSQEQNKLMIEYFEFNI VFKDDKNMVHGFAAER
 KIQAALLDRPALQVDDIGQWSPWTRGEKRIHSCNPNLDIRDQREISQIFLT
 AIGDQARSQRKQSENELMWDQFRKNIAAPNACNNTYQGCTCRGFSOGKKGAOT
 AIALHSLDFPKLDMATVVEKTKVVAHMLFAPESMLVDEGPLSDVGYVMKKNGK
 IYGFGEKDFSFYIHDMEEYKYLKGPVSYOGNVEYFEPQVGRDTMLFYSRIAGV
 PRSLSSQYRIIYLSRNMENMVVIFDLVSTRELRKDLFEVQFMDKCLDIYAR
 LSDQUTISNYSLSNNWLPFINGAAVKNQSDVSDRLQLLAQTLVKEQVAPVM
 RELREALTETPITSLTDVLGILSRKLMKQFANKIAGVGFVGMVTLICFPYKVLIT
 WAKDTPNGELCYENSHKTKIVFLSVVAIGITILMRDRIDGLVKKLDMEIDRFL
 AHVLDVENCPRYYEINDFFSLYSASESGETVLPDLSEVKAQSDKLLQOKKEADEFL
 SAKFSNYSSSVTSPPSVVSGSRSGLLGLEDNSVLTQARVGVSRKDDDEEMEQFL
 SGLIDTEAEIDEVVSFAFSAECERGETSGTKVLCKPLTPPGFENLPAVKPLYSKGTIV
 KRDYFQVGMGRLPKRPVVGSDNSVDAREFLYLDABERVAQNDREIMSLYRDYSRGV
 IRTGQYVPHGLGVGVENKWNWCIRPVVTEHAYVFQDPRMDWDSGLYEVAVWERMGL
 VNSDPAVERMSDVIVGCDOTYLCNNRLIDNLSDALDGPVNCSELPVLDVPGCGKSTMI

mRNA

CDS

Db 1870 ATAGAAATCAAGAGCCACTGCCGAATTCACCAATGCTTAGTCCAGCAGTTTAAACTC 1929

Qy 4837 ggtgagcgtaccattccacgtgtagcagcagaggtgctatccatccatcagtaaacccatc 4896

Db 1930 CAAGCGCCATCCCATCCACGAGATGATCAACCGTGTATCCAAAGGGTCCACAGGTT 1989

Qy 4897 ttacgctcaatcgtggtggaagcaaaccttcgactaagtcaggaagaggtggaag 4956

Db 1990 CTACACATCAACCACTCAGGGGAATGCCTCACCAATAGCTTCCCAAAAGG-----A 2043

Qy 4957 gtcaggtcacaacgtagcttcggtgactattttctatgcttcggttcggttttcaagg 5016

Db 2044 AAGCAAGTATACTAGGATCGGTGACTACTACTAGCCAGGAGTCCAGAG 2103

Qy 5017 acgcaattgattcagtaaacctccatccacgagggcgcatcagttgacgttcag 5072

Db 2104 TCCACAAGCATCGGTAAAGCAACACACACAGGGGAAGGCGTCTCTGACTTTTCAG 2159

RESULT 14

AB045371 9293 bp RNA VRL 03-OCT-2000

LOCUS Apple stem pitting virus gene for RNA polymerase, complete cds.

AB045371

AB045371.1 GI:10566816

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (sites)

Yoshikawa, N., Matsuda, H., Oda, Y., Isoai, M., Takahashi, T., Ito, T. and Yoshida, K.

Genome heterogeneity of apple stem pitting virus in apple trees

Unpublished (2001)

2 (bases 1 to 9293)

Yoshikawa, N.

Direct Submission

Submitted (30-JUN-2000) to the DBI/EMBL/GenBank databases.

Nobuyuki Yoshikawa, Iwate University, Faculty of Agriculture, Ueda

3-18-8, Morioka, Iwate 020-8550, Japan

{E-mail: yoshikawa@iwate-u.ac.jp. Tel: 81-19-621-6150.

Fax: 81-19-621-6150}

Location/Qualifiers

1..9293

/organism="apple stem pitting virus"

/isolate="IF38"

/db_xref="taxon:35350"

61..6618

/gene="ORF1"

61..6618

/gene="ORF1"

/codon_start=1

/product="RNA polymerase"

/protein_id="BAB15950.1"

/db_xref="GI:10566817"

/translation="MALLSRTAAEEVIASTSEEQSRISTQAVLITNVEKKHDLFN

YALPELAKMKSGLYLSHSTPHSHVPCVTKLENNILNPLPSYLDNSFLVLSIKK

NKVDLKRHPDLQWVETINRYSISLDKTRYGFFHVPSPKISAKFKDRRTGFCESDA

SLVDLPFCMEGARKEFFFDHLYWSKEALITFLDHRKPEVMLASIVFPEILAGAK

ESLNPWCYFTIKNDLVFPDGEQSEAVIQPVSGYLLRTGKTTPSPGSDVQLDLK

SSFSHLISITKEAGTGOKRPFESGFEAVSMKGLNPLRKVESCLPISKSTLKIYRY

LRTLKKPDLSAMKLSQVCKDPNGVEIKFPEFSKLCCLKCTLTNTMLDPMRKIVQG

FLFLKPNLRNFVKVQQLHLDNFLETLEEFNFSVGTEDLSLWKDDLEFLNLTGE

TFDVBASSEAWGLKDYVDITVHTTYSVTKEKYDHFQHSLSVDSTLALTKIA

RIVLSYQACTPESSEARVSNLAVNIIAANLACFAVTDLMTTPEKILKEGRKAK

GRMKRFFKFKLQPLTLDVLRVEVETLNSHHNPDDLSEKHAHDLVVPVGNV

MSLRQTFKFKLQPLTLDVLRVEVETLNSHHNPDDLSEKHAHDLVVPVGNV

GCSPVPCNPLCFVQNAVNALHGLNVLDFETGGARGAGSFYSDLAGLSTGTGSH

QSRGWPDFLDQGVENKIPKRYNOCLMVOEYSTGHLSMRDIEDIDYDANHQVLTNY

TGEAFIVDSYEGHSDRATLQKMRKAGKNGRPDNDSEMPAGHAKWIEDSSGA

PVQEFILQIDSLSLEVANKSLSKVDVINCDCMLCNSPLKNEELKFSKALDLAFA

SGFNPTDRSLARGTEGVRSVRIISDLTFHFLRGMHVVDLDDKTTITDIKEGFI

FEATURES

source

gene

CDS

SGPRRKITSCSTDLIMLAFLEPKMTLGGELRAHEDECELSDLTEKHGCSIILSRKF

EPDLFHSFVDEAGNGCFWHSVGLPIGDGEYIKKIILHQAKKLVCKPRLSKLEGNT

WABREAVAYFCSHYGIRLNLVITREECTWIFKHEVLKAAKLVCONHFKPCMPVNC

VIRIASALNGRREVDLAVLGKPAHEDLFEVAEGRGFSIFDLTRFLFISGCSVDA

GGELIMVNGRIRPAEFLSKLEHLAHPVLSRRKFSPIVSDLRNRSAMRFTALANGA

EIDYRPSIDRASTLLDSFELGATGVLCOGKKEAKGLSSMLIPECVDRKLIILCTF

CGKSLFKFKFEKSPGKAITFVSPRSLAESINHDGLAKNGRSGSKDKLNVRV

KTEFLVHLHSDSKEGHTVVIDEIQLPFGYIDLIIILGKQNVNIIAGDPFCSDYDS

SSDRHIFPAGESIDMRILSGRSYKFNILSQFRNPFIYGRPCNLKARUTLDEEYT

LWDSIOFENMKDKDPVVLVSFFSEFKKIVAAHLGLKMKICTYGESITLNFQKADILV

ITDESLTSDRRMTALSRFSDIHFINGMGTWDAITHEVFGPLKHLFFPKRACNDI

EDWFKTHIPIMGLEAVRAQVNDRIILSREVREFRIGITTEQFTDEISKNRGLBETNA

ERYEALPRHKGTDTATFLMAVKKLSFSSPAEAHAKLRARDFGFLDLTFKRVPL

NSRHNESMSEAVHAEFEKLSKSMATIEHNSGRCEDWPIDKALIFMKSQLCCTFDN

FRSAKAGQTLACFOHADSODHFLAFELVEMKFLGLPDLIADYTFIKTHIGSKLGNF

HFPNGVCTESDYEAFOADHFLAFELVEMKFLGLPDLIADYTFIKTHIGSKLGNF

ALMRFTEASTFLFNTMANNLFTFLYDLNGKEAICFAGDDMCANSLKVTNQHARFL

DKIKLAKVQFTTITPTFCGWLCEHGVFKPDLVLERLQIARETNLENICDINIAIEV

SCAYKNGENLNYLTPOEVDHAHNCVRFIVQHNLYLKSIRDIFRGVEVPSVS"

BASE COUNT 2564 a 1856 c 2210 g 2663 t

Query Match 1.08; Score 69.8; DB 79; Length 9293;

Best Local Similarity 52.18; Pred. No. 2.1e-08;

Matches 208; Conservative 0; Mismatches 182; Indels 9; Gaps 2;

Qy 4678 gggaggagtcggtcttttctcaagctgggtgaggggtatccatataatcggtgtagc 4737

Db 2218 GGACGAGTGCCAGCTTTTACTCTAGGGATCTTAAGGGTTACACATTATACAGGATTGAGC 2277

Qy 4738 catgtttcatcaggtggtgctctgctcctagagagatctctaaacggaattaaagtaaccca 4797

Db 2278 CATCAGTCAAGGGGTGCGCAGATTTCTCTGCACAGATTTTGGATGAAATAAGATCCCC 2337

Qy 4798 agcgtctt---caccaactgttttagtcagaagacagaatgaggtgagcgctaccattc 4854

Db 2338 AGAAAGTTTTACACACCACTAGTCTAGTTCAGGAGTACAGCACACTGCTCATGGCTATCAATG 2397

Qy 4855 cagcgtgatcagcagagtagtctatccatcagataaacctctctgacggtcaactctcgtg 4914

Db 2398 CACAGGACGATGAGGATATCTATGATGCTAACCATCAGGTTCTGACTGTCACTATACT 2457

Qy 4915 ggaagagcaaacctctcgactaagtgcaagaggtggtgaaggtcatggtcataaacgta 4974

Db 2458 GGTGAAGCAATTTTTTTCATGATTGCGAGGGCGTAGGT-----TTTGAAGTTAAACTT 2511

Qy 4975 gcttcgggtgactatttcttatgcttcggttttcaaggagcagcacttgactcagta 5034

Db 2512 GATGGCGCCGAAATGCTCTTAATGCCATTTTGGATTTTCAAGAATACAAAGCATGGGATC 2571

Qy 5035 aactccatcagcaagggcgcatcagtttgacgttcagg 5073

Db 2572 AAGTCACCATCTAAGGGAGGAGTTTCATTAACTTTTCGG 2610

RESULT 15

PEBVARN1

LOCUS

DEFINITION

Pea early browning virus (PEBV) genomic RNA1 for 141K, 201K, 30K and 12K proteins.

ACCESSION

X14006

VERSION

X14006.1

KEYWORDS

12 kDa protein; 141 kDa protein.

SOURCE

Pea early browning virus.

ORGANISM

Pea early browning virus

REFERENCE

1 (bases 1 to 7073)

MacFarlane, S.A.

Direct Submission

JOURNAL

Submitted (12-JAN-1989) MacFarlane S.A., John Innes Institute, AFRC Institute of Plant Science Research, Colney Lane, Norwich, NR4 7UH,

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 14, 2001, 15:11:38 ; Search time 216.15 Seconds
(without alignments)
11668.745 Million cell updates/sec

Title: US-09-301-906-3
Perfect score: 6714
Sequence: 1 atggactacattccgccatt.....gttgcttcagggtatatga 6714

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues 960044
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_36.*
1: /SID56/gcgdata/geneseq/geneseqn/NA1980.DAT.*
2: /SID56/gcgdata/geneseq/geneseqn/NA1981.DAT.*
3: /SID56/gcgdata/geneseq/geneseqn/NA1982.DAT.*
4: /SID56/gcgdata/geneseq/geneseqn/NA1983.DAT.*
5: /SID56/gcgdata/geneseq/geneseqn/NA1984.DAT.*
6: /SID56/gcgdata/geneseq/geneseqn/NA1985.DAT.*
7: /SID56/gcgdata/geneseq/geneseqn/NA1986.DAT.*
8: /SID56/gcgdata/geneseq/geneseqn/NA1987.DAT.*
9: /SID56/gcgdata/geneseq/geneseqn/NA1988.DAT.*
10: /SID56/gcgdata/geneseq/geneseqn/NA1989.DAT.*
11: /SID56/gcgdata/geneseq/geneseqn/NA1990.DAT.*
12: /SID56/gcgdata/geneseq/geneseqn/NA1991.DAT.*
13: /SID56/gcgdata/geneseq/geneseqn/NA1992.DAT.*
14: /SID56/gcgdata/geneseq/geneseqn/NA1993.DAT.*
15: /SID56/gcgdata/geneseq/geneseqn/NA1994.DAT.*
16: /SID56/gcgdata/geneseq/geneseqn/NA1995.DAT.*
17: /SID56/gcgdata/geneseq/geneseqn/NA1996.DAT.*
18: /SID56/gcgdata/geneseq/geneseqn/NA1997.DAT.*
19: /SID56/gcgdata/geneseq/geneseqn/NA1998.DAT.*
20: /SID56/gcgdata/geneseq/geneseqn/NA1999.DAT.*
21: /SID56/gcgdata/geneseq/geneseqn/NA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6714	100.0	6714	21	GLRAV-3 ORF 1a, en
2	6714	100.0	17919	21	Grapevine leafroll
3	2557	38.1	4173	18	Grapevine leafroll
4	873	13.0	873	21	GLRAV-3 polyprotei
5	816	12.2	816	21	GLRAV-3 polyprotei
6	360	5.4	360	21	GLRAV-3 polyprotei
7	162.8	2.4	931	21	PMWav-2 ORF 1a DNA
8	162.8	2.4	10000	21	Pineapple mealybug
9	52.8	0.8	8743	20	Rupetris stem pit
10	41	0.6	6741	21	Gene encoding a su
c 11	37.8	0.6	5510	12	Rice actin gene RA
c 12	37.8	0.6	5643	18	T80055

13	36.4	0.5	10732	21	A10594
14	36.2	0.5	1799	21	A08684
15	36.2	0.5	5217	21	A08697
16	35.4	0.5	210	15	O62845
17	35.4	0.5	210	15	O58326
18	35.4	0.5	210	16	Q74917
19	35.4	0.5	210	17	T48084
20	35.4	0.5	210	17	T30992
21	35.4	0.5	210	17	T06736
c 22	35.4	0.5	5059	20	X84332
23	35.4	0.5	7920	20	V08864
24	35.4	0.5	15500	20	V08874
c 25	35.2	0.5	1664976	19	V21209
26	34.8	0.5	1871	13	Q26959
27	34.8	0.5	6384	15	Q71546
c 28	34.6	0.5	1239	14	Q48415
c 29	34.6	0.5	1239	16	T29930
c 30	34.6	0.5	1239	16	T07649
c 31	34.6	0.5	1239	16	T00645
c 32	34.6	0.5	2383	21	A35134
c 33	34.6	0.5	2575	15	Q54641
c 34	34.6	0.5	2575	20	V73321
c 35	34.6	0.5	2988	21	A35135
c 36	34.6	0.5	3365	11	Q05751
c 37	34.6	0.5	6714	21	Z49208
c 38	34.6	0.5	17919	21	Z49200
c 39	34.2	0.5	400	18	V78102
c 40	34.2	0.5	17310	18	V74334
c 41	34.2	0.5	1230025	20	X91990
c 42	34	0.5	324	20	X33948
c 43	34	0.5	565	14	Q35072
c 44	34	0.5	3980	15	Q62696
c 45	33.6	0.5	624	21	A08749

ALIGNMENTS

RESULT 1	
Z49208	
ID	Z49208 standard; DNA; 6714 BP.
XX	
AC	Z49208;
XX	
DF	07-MAR-2000 (first entry)
XX	
DE	GLRAV-3 ORF 1a, encoding polyprotein.
XX	
KW	Grapevine; leafroll; grapevine leafroll-associated virus 3; GLRAV-3; viral disease; yield loss; sugar content; inhibition; infection; replication; polyprotein; domain; proteinase; methyltransferase; helicase; RNA-dependent; RNA polymerase; untranslated region; transgenic plant; component; resistant; truncation; deletion; antisense; expression; detection; antibody; ds.
XX	
OS	Grapevine leafroll-associated virus 3.
XX	
PH	Location/Qualifiers
FT	1..6714
FT	/*tag= a
FT	/note= "ORF 1a"
FT	/product= "GLRAV-3 polyprotein"
mat_peptide	253..612
FT	/*tag= b
FT	/product= "GLRAV-3 polyprotein proteinase domain (Y58143)"
mat_peptide	1378..2193
FT	/*tag= c
FT	/product= "GLRAV-3 polyprotein methyltransferase domain (Y58144)"
mat_peptide	5764..6636
FT	/*tag= d
FT	/product= "GLRAV-3 polyprotein helicase domain (Y58145)"

Db 1381 aagccgcgggaaagagagctgaggaaactctttccagagcttttcgatacacagtctcc 1440
Qy 1441 gactcgcgcggagtagtcacccattcgttaataagcattcgcggagctgtttcaatgaatc 1500
Db 1441 gactcgcgcggagtagtcacccattcgttaataagcattcgcggagctgtttcaatgaatc 1500
Qy 1501 tttccagagaggtgtgtaagtgtgtcttcttcgataattggggagcttcacgtatcat 1560
Db 1501 tttccagagaggtgtgtaagtgtgtcttcttcgataattggggagcttcacgtatcat 1560
Qy 1561 gtcaaaagctggccatgtgaactgtcatgtatgcaatccagctcctagacgttaagaatgtg 1620
Db 1561 gtcaaaagctggccatgtgaactgtcatgtatgcaatccagctcctagacgttaagaatgtg 1620
Qy 1621 aagcgagaaatcaatagatccctcttcttccacagctggggagagattcgtacgtctcc 1680
Db 1621 aagcgagaaatcaatagatccctcttcttccacagctggggagagattcgtacgtctcc 1680
Qy 1681 agtgaccttcaactaaagcgcgttcaaaagtctgtcttcttacttactgtacgtacgaatcgag 1740
Db 1681 agtgaccttcaactaaagcgcgttcaaaagtctgtcttcttacttactgtacgtacgaatcgag 1740
Qy 1741 aactcgcattctagaccgagtcgcgggttttatggtgagtggtacgatatatcccgcgag 1800
Db 1741 aactcgcattctagaccgagtcgcgggttttatggtgagtggtacgatatatcccgcgag 1800
Qy 1801 caggtagcagagctatggaataagaggggtgcgctgggttttcacacatagctcttatgttc 1860
Db 1801 caggtagcagagctatggaataagaggggtgcgctgggttttcacacatagctcttatgttc 1860
Qy 1861 cccgtgaggtgtgtacgttaacgtgaagtttacttggaaacctcgatacgttggtg 1920
Db 1861 cccgtgaggtgtgtacgttaacgtgaagtttacttggaaacctcgatacgttggtg 1920
Qy 1921 aagaggaaggtgattacctcggcctacaattgttggtcagtggtgtgagatgtatgaacat 1980
Db 1921 aagaggaaggtgattacctcggcctacaattgttggtcagtggtgtgagatgtatgaacat 1980
Qy 1981 tcctctctaaagtaagcgggttttcaaccttttcttatgtacgacacttcgctcgggaac 2040
Db 1981 tcctctctaaagtaagcgggttttcaaccttttcttatgtacgacacttcgctcgggaac 2040
Qy 2041 gtgtttaaagctagatgagggataccgtgtgtgtgtacacatctcactatgtatagg 2100
Db 2041 gtgtttaaagctagatgagggataccgtgtgtgtgtacacatctcactatgtatagg 2100
Qy 2101 gctcagaagtcacactggaactgaggttacgtataggtgtgtgtcccgctgttcgtgggc 2160
Db 2101 gctcagaagtcacactggaactgaggttacgtataggtgtgtgtcccgctgttcgtgggc 2160
Qy 2161 aaatcgtcgtgttcatacctgtgttagctgtgttcagtggtgttcctttaagacaatagtc 2220
Db 2161 aaatcgtcgtgttcatacctgtgttagctgtgttcagtggtgttcctttaagacaatagtc 2220
Qy 2221 ctgcattcggactttctcacagagatctatctcagcgcctcaacactataggacattc 2280
Db 2221 ctgcattcggactttctcacagagatctatctcagcgcctcaacactataggacattc 2280
Qy 2281 gagaatagaacgtttgagtatgcgtttggggttcaggttcgcaaaagacccatgtcatt 2340
Db 2281 gagaatagaacgtttgagtatgcgtttggggttcaggttcgcaaaagacccatgtcatt 2340
Qy 2341 acagggagtcgctgtgtccacagcaaggttgatatttctcctgatgatgtggggttta 2400
Db 2341 acagggagtcgctgtgtccacagcaaggttgatatttctcctgatgatgtggggttta 2400
Qy 2401 gttgtcgcgttatgtctcagcaggttaaggaataggcgaagatttcgctccttaaac 2460
Db 2401 gttgtcgcgttatgtctcagcaggttaaggaataggcgaagatttcgctccttaaac 2460
Qy 2461 ttataaaagccagtgagggagctcgcggggtcttcaagctctcttccagaccgta 2520
Db 2461 ttataaaagccagtgagggagctcgcggggtcttcaagctctcttccagaccgta 2520

Qy 2521 ggcgattgtttttcgaacgcagtcctcgtctatgctaaagcaatggtgcacgataaacttc 2580
Db 2521 ggcgattgtttttcgaacgcagtcctcgtctatgctaaagcaatggtgcacgataaacttc 2580
Qy 2581 aacgtttttgagagcgtttatgtctatgcccagagcgttccatccgtaaagtaacgttggtct 2640
Db 2581 aacgtttttgagagcgtttatgtctatgcccagagcgttccatccgtaaagtaacgttggtct 2640
Qy 2641 gttgtgttaccatttgcaactcttgagcttcagacaggttgagctcaggggtgccttt 2700
Db 2641 gttgtgttaccatttgcaactcttgagcttcagacaggttgagctcaggggtgccttt 2700
Qy 2701 gatatttcgaagagaccccttcgtaggaactgaagaatagtcgcttgcgctctctct 2760
Db 2701 gatatttcgaagagaccccttcgtaggaactgaagaatagtcgcttgcgctctctct 2760
Qy 2761 agggctatcgtggaagattcaattaaaggtcatgaagcaatgaagacagaatggaaaa 2820
Db 2761 agggctatcgtggaagattcaattaaaggtcatgaagcaatgaagacagaatggaaaa 2820
Qy 2821 cccctgccaattactgaagattctgtatatcggttcaataatggggaacgtttctcaacgtc 2880
Db 2821 cccctgccaattactgaagattctgtatatcggttcaataatggggaacgtttctcaacgtc 2880
Qy 2881 cactgtacgagggcaggtctcttgcggttcgaaaaagccgttggtttcagtggtttct 2940
Db 2881 cactgtacgagggcaggtctcttgcggttcgaaaaagccgttggtttcagtggtttct 2940
Qy 2941 aaggggttctgtagctcgtgggctgcagcaagcgtttcttgccattacgtctcttt 3000
Db 2941 aaggggttctgtagctcgtgggctgcagcaagcgtttcttgccattacgtctcttt 3000
Qy 3001 tccacaggttccactattctacgaccgcgttttaactgaagatgaagcgtttgaagctcgt 3060
Db 3001 tccacaggttccactattctacgaccgcgttttaactgaagatgaagcgtttgaagctcgt 3060
Qy 3061 gtgcgacagagaatgctataaactcaacgggttgggcactgtagagacgttcgcgtagct 3120
Db 3061 gtgcgacagagaatgctataaactcaacgggttgggcactgtagagacgttcgcgtagct 3120
Qy 3121 gtgagcaaggtcgtagctggaacgaagaattttggaagtaagtttcttcttaaaagacttc 3180
Db 3121 gtgagcaaggtcgtagctggaacgaagaattttggaagtaagtttcttcttaaaagacttc 3180
Qy 3181 accactttcgtatttcggaataaaggtctcttatcgggataattcgtggtctttgggtgctg 3240
Db 3181 accactttcgtatttcggaataaaggtctcttatcgggataattcgtggtctttgggtgctg 3240
Qy 3241 gccccaattgcatgggaagtataggcgggaatttcgggtaacgtagaaggtacgcgggc 3300
Db 3241 gccccaattgcatgggaagtataggcgggaatttcgggtaacgtagaaggtacgcgggc 3300
Qy 3301 agtagttacgaactctaaagctcgttaagtctcaaacgcccggttggtttacgcgggttta 3360
Db 3301 agtagttacgaactctaaagctcgttaagtctcaaacgcccggttggtttacgcgggttta 3360
Qy 3361 acctctagcacagtatccggtggtatcttttagtcgtgcgaagaggtttctcgtcgcggctg 3420
Db 3361 acctctagcacagtatccggtggtatcttttagtcgtgcgaagaggtttctcgtcgcggctg 3420
Qy 3421 accgctacagggcagccgttagctaaacgtcaagtcaccttagcgttgctatgctgtttct 3480
Db 3421 accgctacagggcagccgttagctaaacgtcaagtcaccttagcgttgctatgctgtttct 3480
Qy 3481 acctcaacgcatttccggctcagtagtttaggcattttgggcacatgctcttccacgg 3540
Db 3481 acctcaacgcatttccggctcagtagtttaggcattttgggcacatgctcttccacgg 3540
Qy 3541 cacttaagtgtttcttctgttttagggacatttctggggtcaggggttagcgcgaatact 3600
Db 3541 cacttaagtgtttcttctgttttagggacatttctggggtcaggggttagcgcgaatact 3600

Db 5761 atggtggcgcaatcgcttaagtcttcgaataataagtgctataatgctccccaggtggc 5820
QY 5821 ggtaaagcagcgcgttagtgacgaattcgttaagtcacccaatagcacagccacatt 5880
Db 5821 ggtaaagcagcgcgttagtgacgaattcgttaagtcacccaatagcacagccacatt 5880
QY 5881 acggtcaatgtgggaagttctgagcacataaataatgctggtgaagaagagatccgaat 5940
Db 5881 acggtcaatgtgggaagttctgagcacataaataatgctggtgaagaagagatccgaat 5940
QY 5941 ttggaaggtctcaacagtgctacacagtttaactccaggtggttaaaactttatcgtcagg 6000
Db 5941 ttggaaggtctcaacagtgctacacagtttaactccaggtggttaaaactttatcgtcagg 6000
QY 6001 ggaatgtataaaaggtttttgtgtagtgaggtgcacatgatgcatacaagcttactacaa 6060
Db 6001 ggaatgtataaaaggtttttgtgtagtgaggtgcacatgatgcatacaagcttactacaa 6060
QY 6061 ctaggcgtcttcgaaccgcgcgtcggaagcctcttttttgagacataaatacagata 6120
Db 6061 ctaggcgtcttcgaaccgcgcgtcggaagcctcttttttgagacataaatacagata 6120
QY 6121 ccaattcataaacaggaggaaggtgttttagtgattgtgctgtttttgttccaaagaag 6180
Db 6121 ccaattcataaacaggaggaaggtgttttagtgattgtgctgtttttgttccaaagaag 6180
QY 6181 gaaagcgttgtatacaccttctaactcgaaggtgctgctgtagtgattgtgctgtttttgttccaaagaag 6240
Db 6181 gaaagcgttgtatacaccttctaactcgaaggtgctgctgtagtgattgtgctgtttttgttccaaagaag 6240
QY 6241 tccctcaatcagcgttaagggaacggaaggtgttaccctgaaaggtcgttagcgttaag 6300
Db 6241 tccctcaatcagcgttaagggaacggaaggtgttaccctgaaaggtcgttagcgttaag 6300
QY 6301 gacaaaccagtagtagatgcgtgtccaaagggccaaattggaaccactgatgcgttagct 6360
Db 6301 gacaaaccagtagtagatgcgtgtccaaagggccaaattggaaccactgatgcgttagct 6360
QY 6361 gaaataaacgctgacgtgtacttctgtagtcacaggttggaagtcgagatagaagagg 6420
Db 6361 gaaataaacgctgacgtgtacttctgtagtcacaggttggaagtcgagatagaagagg 6420
QY 6421 tcggtgaagggaaggaagaaacacacagtgatgacagtgatgacagtgacaggaagaa 6480
Db 6421 tcggtgaagggaaggaagaaacacacagtgatgacagtgatgacagtgacaggaagaa 6480
QY 6481 acattcagtgatgtgtattgttttagacgaagaagccgagtgactccctattcactaaa 6540
Db 6481 acattcagtgatgtgtattgttttagacgaagaagccgagtgactccctattcactaaa 6540
QY 6541 caaccgcataacttctgtgtttgttcgagacacacacgctcactggtttatgcccgtctg 6600
Db 6541 caaccgcataacttctgtgtttgttcgagacacacacgctcactggtttatgcccgtctg 6600
QY 6601 agctcaaatgttgacgataaggttcgacacataatattagcagcgtcactcaatcagta 6660
Db 6601 agctcaaatgttgacgataaggttcgacacataatattagcagcgtcactcaatcagta 6660
QY 6661 tccgagcgtttgtctcacagcttcgcccggtggtgtgttttcgaggttatga 6714
Db 6661 tccgagcgtttgtctcacagcttcgcccggtggtgtgttttcgaggttatga 6714

RESULT 2

Z49200

ID Z49200 standard; DNA; 17919 BP.

XX

AC Z49200;

XX

DT 07-MAR-2000 (first entry)

XX

DE Grapevine leafroll-associated virus 3 (GLRaV-3) genome.

XX

29-APR-1999; 99WO-US09307.

KW Grapevine; leafroll; grapevine leafroll-associated virus 3; GLRaV-3;
viral disease; yield loss; sugar content; inhibition; infection;
KW replication; polypeptide; domain; proteinase; methyltransferase;
KW helicase; RNA-dependent; RNA polymerase; untranslated region;
KW transgenic plant; component; resistant; truncation; deletion; antisense;
expression; detection; antibody; ds.
XX
OS Grapevine leafroll-associated virus 3.
XX
FH Key Location/Qualifiers
FT 5'UTR 1..158
FT FT /*tag= a
FT CDS 159..6872
FT FT /*tag= b
FT FT /note= "ORF 1a"
FT FT /product= "GLRaV-3 polyprotein (Y58148)"
FT mat_peptide 411..770
FT FT /*tag= c
FT FT /product= "GLRaV-3 polyprotein proteinase domain
(Y58143)"
FT mat_peptide 1536..2351
FT FT /*tag= d
FT FT /product= "GLRaV-3 polyprotein methyltransferase domain
(Y58144)"
FT mat_peptide 5922..6794
FT FT /*tag= e
FT FT /product= "GLRaV-3 polyprotein helicase domain (Y58145)"
FT CDS 6877..8478
FT FT /*tag= f
FT FT /*tag= g
FT FT /note= "ORF 1b"
FT FT /product= "GLRaV-3 RNA-dependent RNA polymerase (Y58146)"
FT CDS 8708..8863
FT FT /*tag= h
FT FT /note= "ORF 2"
FT CDS 9930..10067
FT FT /*tag= i
FT FT /note= "ORF 3"
FT CDS 10086..11735
FT FT /*tag= j
FT FT /note= "ORF 4"
FT CDS 11728..13179
FT FT /*tag= k
FT FT /note= "ORF 5"
FT CDS 13269..14210
FT FT /*tag= l
FT FT /note= "ORF 6"
FT CDS 14273..15706
FT FT /*tag= m
FT FT /note= "ORF 7"
FT CDS 15717..16274
FT FT /*tag= n
FT FT /note= "ORF 8"
FT CDS 16271..16804
FT FT /*tag= o
FT FT /note= "ORF 9"
FT CDS 16811..17350
FT FT /*tag= p
FT FT /note= "ORF 10"
FT CDS 17353..17463
FT FT /*tag= q
FT FT /note= "ORF 11"
FT FT /product= "GLRaV-3 ORF 11-encoded protein (Y58147)"
FT CDS 17460..17642
FT FT /*tag= r
FT FT /note= "ORF 12"
FT 3'UTR 17643..17919
FT FT /*tag= r
XX
PN WO9955880-A1.
XX
PD 04-NOV-1999.
XX
PF 29-APR-1999; 99WO-US09307.

Db 1659 ttttccagggggtgtgtaactgtgtcttcttcgatatgtggggggagctccagtgatcat 1718
QY 1561 gtcaagctggccatgtgaactgtcatgtatgcaatccagctcctagacgtttaaagatgtg 1620
Db 1719 gtcaagctggccatgtgaactgtcatgtatgcaatccagctcctagacgtttaaagatgtg 1778
QY 1621 aagcgagagaataaatgagatcctcttcttccacagctggggagattcgtacgtctcc 1680
Db 1779 aagcgagagaataaatgagatcctcttcttccacagctggggagattcgtacgtctcc 1838
QY 1681 agtgacctctaaactgaagcggctcaagctctgtcttactactgtactgtactgcaagatccag 1740
Db 1839 agtgacctctaaactgaagcggctcaagctctgtcttactactgtactgtactgcaagatccag 1898
QY 1741 aactcgattctagagccgatgcgggtttttatgtgtgattgtgtacgatataatcccgcag 1800
Db 1899 aactcgattctagagccgatgcgggtttttatgtgtgattgtgtacgatataatcccgcag 1958
QY 1801 caggtagcagaggtctatgataagaagggtgcgctggttttccacatagctcttattgttc 1860
Db 1959 caggtagcagaggtctatgataagaagggtgcgctggttttccacatagctcttattgttc 2018
QY 1861 cccgtgaggtgtgtacggttaacggtgaagtttaacttggaaagaaactcgatacgtttgggtg 1920
Db 2019 cccgtgaggtgtgtacggttaacggtgaagtttaacttggaaagaaactcgatacgtttgggtg 2078
QY 1921 aagagggaaaggtgatcacctggcctacaattgttggtcagtggtgtgagatgtatgaacat 1980
Db 2079 aagagggaaaggtgatcacctggcctacaattgttggtcagtggtgtgagatgtatgaacat 2138
QY 1981 tccttctcaacgtaagcgggtttttccacgttttcttatagtacgcacttcgtccgggaac 2040
Db 2139 tccttctcaacgtaagcgggtttttccacgttttcttatagtacgcacttcgtccgggaac 2198
QY 2041 gtgtttaaactgaagatagaggataccgttgggttaccatactcactatgtatagg 2100
Db 2199 gtgtttaaactgaagatagaggataccgttgggttaccatactcactatgtatagg 2258
QY 2101 gtcagaagtcacctggaactgaggttaagtaagtcgttgggtcccgctcgttcgtgggc 2160
Db 2259 gtcagaagtcacctggaactgaggttaagtaagtcgttgggtcccgctcgttcgtgggc 2318
QY 2161 aaatcgctgggttcatacctgtgtagctgtgtctagtggttccttttaagacaaatagtc 2220
Db 2319 aaatcgctgggttcatacctgtgtagctgtgttcctagtggttccttttaagacaaatagtc 2378
QY 2221 ctcgattcgacctgtgcacaggaatctattcctacgcgtcaacactataggacattc 2280
Db 2379 ctcgattcgacctgtgcacaggaatctattcctacgcgtcaacactataggacattc 2438
QY 2281 gagaaatagaacgtttgagtatgcgttggggcggtcaggtcgcgaagaccatgtcatt 2340
Db 2439 gagaaatagaacgtttgagtatgcgttggggcggtcaggtcgcgaagaccatgtcatt 2498
QY 2341 acagggagtcggtgttccacagcaagggtgatatttctcctgatgatgtgggggttta 2400
Db 2499 acagggagtcggtgttccacagcaagggtgatatttctcctgatgatgtgggggttta 2558
QY 2401 gttgtcgctgttatggtctcaggcgttaaggataggcggaagagatttcgctcctataac 2460
Db 2559 gttgtcgctgttatggtctcaggcgttaaggataggcggaagagatttcgctcctataac 2618
QY 2461 tttataaaagccagtgaggggagtcgcgggggtcttcaagctcttcttcagaccgta 2520
Db 2619 tttataaaagccagtgaggggagtcgcgggggtcttcaagctcttcttcagaccgta 2678
QY 2521 ggogattgttttcgaacagcagtcctcgtctatgctaaaggcaatggtgcacgataacttc 2580
Db 2679 ggogattgttttcgaacagcagtcctcgtctatgctaaaggcaatggtgcacgataacttc 2738
QY 2581 aacgttttggagacgcttatgtctatgcccagagcgttccatccgttaaaagtacctgggtct 2640
|||||

Db 2739 aacgttttggagacgcttatgtctatgcccagagcgttccatccgttaaaagtacctgggtct 2798
QY 2641 gttgtgtttaccatttgcaacttctgagcttcagacaggttgagctcaggggtgccttt 2700
Db 2799 gttgtgtttaccatttgcaacttctgagcttcagacaggttgagctcaggggtgccttt 2858
QY 2701 gatatttcgaagagacaccttcggtagtagaaactgaagaatagtcgctgcgctctctct 2760
Db 2859 gatatttcgaagagacaccttcggtagtagaaactgaagaatagtcgctgcgctctctct 2918
QY 2761 agggctatcgttgaaatccaatttaaggctcatgaaggcaatgaagacagagaatggaaaa 2820
Db 2919 agggctatcgttgaaatccaatttaaggctcatgaaggcaatgaagacagagaatggaaaa 2978
QY 2821 cccctgccaattactgaagattctgtatatgcgttcaataatggggaacgtttcttaacgtc 2880
Db 2979 cccctgccaattactgaagattctgtatatgcgttcaataatggggaacgtttcttaacgtc 3038
QY 2881 cactgtacgagggcagggtcttcttgccggttcgaaagcgaccgtggtttcgagtggtttct 2940
Db 3039 cactgtacgagggcagggtcttcttgccggttcgaaagcgaccgtggtttcgagtggtttct 3098
QY 2941 aagggtttggttagctcgtgggggtgcgacgaagccctttctggcattacgctcgtctctt 3000
Db 3099 aagggtttggttagctcgtgggggtgcgacgaagccctttctggcattacgctcgtctctt 3158
QY 3001 tccacaggttcacattcttacgaccgcggttttaactgaagatgaagagcgttgatgcctcg 3060
Db 3159 tccacaggttcacattcttacgaccgcggttttaactgaagatgaagagcgttgatgcctcg 3218
QY 3061 gtgcgcacagagaatgctataaaactcaccggtggggcatactggagacgtgcgcgtagct 3120
Db 3219 gtgcgcacagagaatgctataaaactcaccggtggggcatactggagacgtgcgcgtagct 3278
QY 3121 gtgaacaaaggtctagctgaacgaagaatatttgagtgaaatttcccttaaatgaattc 3180
Db 3279 gtgagcaaggtctagctggagcaagaagaattttgagtgaaatttcccttaaatgaattc 3338
QY 3181 accacttcgctatttcggaataaagggtgcttattcgggatacttcgtggcgtctttgggtgcg 3240
Db 3339 accacttcgctatttcggaataaagggtgcttattcgggatacttcgtggcgtctttgggtgcg 3398
QY 3241 gccccaattgcatggaagtataggcgcggaatttcgggctaactagaaggttaocggggc 3300
Db 3399 gccccaattgcatggaagtataggcgcggaatttcgggctaactagaaggttaocggggc 3458
QY 3301 agtagttacgaactctaaagctcttaagttcaagtcacagcccgctggtttacgcggttta 3360
Db 3459 agtagttacgaactctaaagctcttaagttcaagtcacagcccgctggtttacgcggttta 3518
QY 3361 accttagcacagtatccgggtggtatttagtcgtcggaagaggggttttcgctcgccggtg 3420
Db 3519 accttagcacagtatccgggtggtatttagtcgtcggaagaggggttttcgctcgccggtg 3578
QY 3421 accgtcacttagggcgaccgttagctaaacgttcaagtcocctcttagcgttgctatgctttct 3480
Db 3579 accgtcacttagggcgaccgttagctaaacgttcaagtcocctcttagcgttgctatgctttct 3638
QY 3481 acctcaacgccaatttcgggctgcagtatgttaggcattttgggcacatgctcttccacgg 3540
Db 3639 acctcaacgccaatttcgggctgcagtatgttaggcattttgggcacatgctcttccacgg 3698
QY 3541 cacttaattgttttcttcttgggttagggacatttagtcgggcgagggctagcgcgaatact 3600
Db 3699 cacttaattgttttcttcttgggttagggacatttagtcgggcgagggctagcgcgaatact 3758
QY 3601 tggaaagtttggagggtcttcccaataattggtgcgtgttcccagaggttgtttggcgagg 3660
Db 3759 tggaaagtttggagggtcttcccaataattggtgcgtgttcccagaggttgtttggcgagg 3818
QY 3661 aagagtgtagcgtcatgttactgcctattacgctaggggtatctttgatcataaaggggc 3720
Db 3819 aagagtgtagcgtcatgttactgcctattacgctaggggtatctttgatcataaaggggc 3878

QY 3721 ttgcttaacgacaccatacctaacttgcttacgttccaccgtagagggagggaatgtg 3780
 Db 3879 tigtctlaacgacaccatacctaacttgcttacgttccaccgtagagggagggaatgtg 3938
 QY 3781 tacgtagaacgcttaggtattaccgggactttgactatgacgaaggtgcttgctccatct 3840
 Db 3939 tacgtagaacgcttaggtattaccgggactttgactatgacgaaggtgcttgctccatct 3998
 QY 3841 gggactcagatgaagcgttcccgtgacgataacgatlgaatccacttctagtctca 3900
 Db 3999 gggactcagatgaagcgttcccgtgacgataacgatlgaatccacttctagtctca 4058
 QY 3901 agctatgattgtgtcaaaatgtgcgcgactgggattagcaccacggggaagttaact 3960
 Db 4059 agctatgattgtgtcaaaatgtgcgcgactgggattagcaccacggggaagttaact 4118
 QY 3961 ggtgaagaagaagaccatcactctgaagcgtgacgtgcaatacacttatgtcaggaagaggtt 4020
 Db 4119 ggtgaagaagaagaccatcactctgaagcgtgcaatacacttatgtcaggaagaggtt 4178
 QY 4021 gcccgctcgcagctgtgcggaagaacaggtgatlccgtcgggttctgtgtaccgctgac 4080
 Db 4179 gcccgctcgcagctgtgcggaagaacaggtgatlccgtcgggttctgtgtaccgctgac 4238
 QY 4081 gctatggctttgttgaagtgtaaaaaaggtgtgcgacgatgtctttccaccaacagtct 4140
 Db 4239 gctatggctttgttgaagtgtaaaaaaggtgtgcgacgatgtctttccaccaacagtct 4298
 QY 4141 agtggggaaacggctcgtgaggttgaggtggaagcgaaggggtgtgtccccaagaagcgtc 4200
 Db 4299 agtggggaaacggctcgtgaggttgaggtggaagcgaaggggtgtgtccccaagaagcgtc 4358
 QY 4201 gtcggtgagcgcacacaaaggggaagcgtgcagatggttaaacacacacaaacc 4260
 Db 4359 gtcggtgagcgcacacaaaggggaagcgtgcagatggttaaacacacacaaacc 4418
 QY 4261 ggggtcaacgaaggcgacaggagccagtcacagtcacgtcttctgtgagttcgccacaggct 4320
 Db 4419 ggggtcaacgaaggcgacaggagccagtcacagtcacgtcttctgtgagttcgccacaggct 4478
 QY 4321 gatattccaaaggtccaccagtcgaggttacatgctcgaagaagagtgaaacaagaagta 4380
 Db 4479 gatattccaaaggtccaccagtcgaggttacatgctcgaagaagagtgaaacaagaagta 4538
 QY 4381 ccaattggcactgtttcgggcgccacgccaatcgtcgtatgagaacccgccccaagtgctt 4440
 Db 4539 ccaattggcactgtttcgggcgccacgccaatcgtcgtatgagaacccgccccaagtgctt 4598
 QY 4441 acgactcgttggttgaagataattgacaaggggcaagggcgctcgtcatgtggtctgagaaa 4500
 Db 4599 acgactcgttggttgaagataattgacaaggggcaagggcgctcgtcatgtggtctgagaaa 4658
 QY 4501 aaacaggtacaagtcgagcagcccaaaacagagaggtttgacgatacaatgaaggcaaggcc 4560
 Db 4659 aaacaggtacaagtcgagcagcccaaaacagagaggtttgacgatacaatgaaggcaaggcc 4718
 QY 4561 ggtlaaacagctttgcatgtttagaacgtgttctcgtcgggtgacgtggtatgtgtacaac 4620
 Db 4719 ggtlaaacagctttgcatgtttagaacgtgttctcgtcgggtgacgtggtatgtgtacaac 4778
 QY 4621 gaagcactatcgccaccagggtttctcaaacgcatattacctttgtcgataacttgaaagg 4680
 Db 4779 gaagcactatcgccaccagggtttctcaaacgcatattacctttgtcgataacttgaaagg 4838
 QY 4681 aggagtcgggtcttttctcaaacgtggtgaggggtatacctataaattggttggtgacct 4740
 Db 4839 aggagtcgggtcttttctcaaacgtggtgaggggtatacctataaattggttggtgacct 4898
 QY 4741 gtttcaatcagggtggcctcgtgccctagaggatatacttaacggcaatgaagtaccacaagc 4800
 Db 4899 gtttcaatcagggtggcctcgtgccctagaggatatacttaacggcaatgaagtaccacaagc 4958

QY 4801 gtcttcgacacactgttttagtcagaaagtacaagatgggtggagcgctaacatttccacgct 4860
 Db 4959 gtcttcgacacactgttttagtcagaaagtacaagatgggtggagcgctaacatttccacgct 5018
 QY 4861 gatgacgaaggagtgtctatccatcaagataaaccctattctgacggctcaatctcgtggggaag 4920
 Db 5019 gatgacgaaggagtgtctatccatcaagataaaccctattctgacggctcaatctcgtggggaag 5078
 QY 4921 gcaaaacttctcactaaagtgcaggaaggttgtaaggtcatggtcataaacgtagcttcg 4980
 Db 5079 gcaaaacttctcactaaagtgcaggaaggttgtaaggtcatggtcataaacgtagcttcg 5138
 QY 4981 ggtgactattttcttatgcttcgcttcttcaaaagcagcacttgcattcagtaaaactcc 5040
 Db 5139 ggtgactattttcttatgcttcgcttcttcaaaagcagcacttgcattcagtaaaactcc 5198
 QY 5041 atcgacgaaggcgcaatcagtttgacgttcaagggaactcggcgctctttgggtgtaggc 5100
 Db 5199 atcgacgaaggcgcaatcagtttgacgttcaagggaactcggcgctctttgggtgtaggc 5258
 QY 5101 aggatgttcagttagccggtgcggtgtcggatgagaagtcaccaggtgttccaaaccag 5160
 Db 5259 aggatgttcagttagccggtgcggtgtcggatgagaagtcaccaggtgttccaaaccag 5318
 QY 5161 caaccaagagccaaaggtgtctaccagaacaaatcacaccaaaatcgggggcaaggctcta 5220
 Db 5319 caaccaagagccaaaggtgtctaccagaacaaatcacaccaaaatcgggggcaaggctcta 5378
 QY 5221 tctgaggggaagtgtgtgaaggaaagtcgaagggaagtcgacataactcgatctggtgcgaacaa 5280
 Db 5379 tctgaggggaagtgtgtgaaggaaagtcgaagggaagtcgacataactcgatctggtgcgaacaa 5438
 QY 5281 gattacgtttagaagtgtagtgggtcagggctgataaaccagtgatgctcttgaacct 5340
 Db 5439 gattacgtttagaagtgtagtgggtcagggctgataaaccagtgatgctcttgaacct 5498
 QY 5341 gactacaccccaatgacatttgaagtgggttaaaacccgggacacctctgaaagtgcgctg 5400
 Db 5499 gactacaccccaatgacatttgaagtgggttaaaacccgggacacctctgaaagtgcgctg 5558
 QY 5401 gagtacttgaagtattcgtctataggtcattgagaggacatacagaaggcgttctgcttggct 5460
 Db 5559 gagtacttgaagtattcgtctataggtcattgagaggacatacagaaggcgttctgcttggct 5618
 QY 5461 agaaaaattgcgctcactaccgccgaaggtgttctgaaagtacctaatcaagtttatgaa 5520
 Db 5619 agaaaaattgcgctcactaccgccgaaggtgttctgaaagtacctaatcaagtttatgaa 5678
 QY 5521 tcactacccgggtttcacggtttacaagtcgggcacagatctcaatttttcatcaacacaa 5580
 Db 5679 tcactacccgggtttcacggtttacaagtcgggcacagatctcaatttttcatcaacacaa 5738
 QY 5581 gacggcttcggtgtgagagacctaccgtacgtactatagctgaaaaaggtatctttacc 5640
 Db 5739 gacggcttcggtgtgagagacctaccgtacgtactatagctgaaaaaggtatctttacc 5798
 QY 5641 aaggggcaaaagtgtcgacgcggtgtgtgttgggcgacaactctgtctgtatgcgacgat 5700
 Db 5799 aaggggcaaaagtgtcgacgcggtgtgtgttgggcgacaactctgtctgtatgcgacgat 5858
 QY 5701 atactggttttccacgatgccatatttgaagtaggtgcactgaaagtcogctcgatcgcc 5760
 Db 5859 atactggttttccacgatgccatatttgaagtaggtgcactgaaagtcogctcgatcgcc 5918
 QY 5761 algggtggcgcaatcggttgaagtccttcgaataataagtgctataaactctccccaggctg 5820
 Db 5919 atgggtggcgcaatcggttgaagtccttcgaataataagtgctataaactctccccaggctg 5978
 QY 5821 ggtlaaacgacacagcgttagtgacgaatctgttaagtcacccaatgacgacgacacatt 5880
 Db 5979 ggtlaaacgacacagcgttagtgacgaatctgttaagtcacccaatgacgacgacacatt 6038
 QY 5881 acggcttaatgtgggaagttctgaggacataaataatgtgggtgaagaagagatccgaat 5940


```
Db 6039 acggctaattggaagttcttgaggacataaataatgagcgtggaagaagagatccgaat 6098
Qy 5941 ttggaaggtctcaacagtgctaccacagtttaactccagggttggtaaaactttatcgtcagg 6000
Db 6099 ttggaaggtctcaacagtgctaccacagtttaactccagggttggtaaaactttatcgtcagg 6158
Qy 6001 ggaatgtataaaagggttttgggtgagtgaggtgcacatgatgatcaaggcttactacaa 6060
Db 6159 ggaatgtataaaagggttttgggtgagtgaggtgcacatgatgatcaaggcttactacaa 6218
Qy 6061 ctaggcgctctcgcaaccgcgctgcggaaggcctcttttttgagacataaaatcacagata 6120
Db 6219 ctaggcgctctcgcaaccgcgctgcggaaggcctcttttttgagacataaaatcacagata 6278
Qy 6121 ccattcataaacagggaagggtgttttaggatgattgtgctgttttttgcctcaaaagaag 6180
Db 6279 ccattcataaacagggaagggtgttttaggatgattgtgctgttttttgcctcaaaagaag 6338
Qy 6181 gaaagcgttgatacacttctaaatcgtaacaggtgcggttagatgattgttctactctgttg 6240
Db 6339 gaaagcgttgatacacttctaaatcgtaacaggtgcggttagatgattgttctactctgttg 6398
Qy 6241 tctcaatgacgttaagggaacggaaggtgttacctgaaaggtcgcttagcgttaag 6300
Db 6399 tctcaatgacgttaagggaacggaaggtgttacctgaaaggtcgcttagcgttaag 6458
Qy 6301 gacaaaccagtagtaagatcgctgtccaaaaggccaaattggaaccactgatgacgtagct 6360
Db 6459 gacaaaccagtagtaagatcgctgtccaaaaggccaaattggaaccactgatgacgtagct 6518
Qy 6361 gaaataaacgtgacgtgactgtgcatgacccagttgagaaagtcggatatgaagagg 6420
Db 6519 gaaataaacgtgacgtgactgtgcatgacccagttgagaaagtcggatatgaagagg 6578
Qy 6421 tcgftgaagggaagaaagaaacaccagtgatgacagtgcatgaaagcacaggaagaaa 6480
Db 6579 tcgftgaagggaagaaagaaacaccagtgatgacagtgcatgaaagcacaggaagaaa 6538
Qy 6481 acattcagtgatggtgtattgttagacgaagaaagccgatgactccctattcactaaa 6540
Db 6639 acattcagtgatggtgtattgttagacgaagaaagccgatgactccctattcactaaa 6598
Qy 6541 caaccgcataactgtgtgtgttcgagacacacacgcgtcaactggtttatgcgcgtctg 6600
Db 6699 caaccgcataactgtgtgtgttcgagacacacacgcgtcaactggtttatgcgcgtctg 6758
Qy 6601 agctcaaatgtgacgataaggctcgacacatatattagcgacgcgtcaactcaatcagta 6660
Db 6759 agctcaaatgtgacgataaggctcgacacatatattagcgacgcgtcaactcaatcagta 6818
Qy 6661 tcgcagcgtttgttcacacgcttcgccccgctggtgtgttttcgaggtatatga 6714
Db 6819 tcgcagcgtttgttcacacgcttcgccccgctggtgtgttttcgaggtatatga 6872

RESULT 3
T72214
ID T72214 standard; cDNA; 4173 BP.
XX AC T72214;
XX DT 19-SEP-1997 (first entry)
XX DE Grapevine leafroll virus helicase cDNA.
XX KW GLRaV; grapevine; vitis; rootstock; leafroll; disease resistance;
XX KW transgenic plant; tristeza virus; citrus; helicase; ds.
XX OS Grapevine leafroll associated virus type 3 isolate NY1.
XX PN WO9722700-A2.
XX
```

```
PD 26-JUN-1997.
XX
PF 20-DEC-1996; 96WO-US20747.
XX
PR 21-DEC-1995; 95US-0009008.
XX
PA (CORR ) CORNELL RES FOUND INC.
XX
PI Gonsalves D, Ling K;
XX
WP; 1997-341691/31.
DR P-PSDB; W21636.
XX
XX DNA encoding grape-vine leaf-roll virus proteins - useful to impart
PT viral-resistance to Vitis scion or root-stock cultivar(s)
PS
XX Claim 5; Page 66-69; 172pp; English.
XX
CC A double-stranded cDNA molecule (T72214) comprises the incomplete
CC coding sequence of the helicase (W21636) of grapevine leafroll
CC associated virus type 3 (GLRaV-3). It was identified as open
CC reading frame 1a in a 15,227-nucleotide sequence that covers about
CC 80% of the GLRaV-3 genome. This sequence was isolated from a cDNA
CC library prepd. from GLRaV-3 dsRNA obtd. from leafroll-diseased
CC grapevine canes. Isolated GLRaV-3 nucleic acids (T72214-25) can be
CC used to prepare GLRaV-3 polypeptides (W21636-47), to impart GLRaV
CC resistance to Vitis scion or rootstock cultivars or tristeza virus
CC resistance to citrus scion or rootstock cultivars, and to design
CC probes for detection of GLRaV.
XX
SQ Sequence 4173 BP; 1030 A; 874 C; 1177 G; 1092 T; 0 other;

Query Match 38.1%; Score 2557; DB 18; Length 4173;
Best Local Similarity 76.3%; Pred. No. 0;
Matches 3187; Conservative 0; Mismatches 980; Indels 12; Gaps 3;

Qy 2542 gtctcgtctatgctaaggcaatggtgcacgataaacttcaacgcttttggagacgcttatg 2601
Db 1 gtgtctacttcacggaagtgatgacgacaatttcaatctccttgagaccttgta 60
Qy 2602 tctatgccagagcgttcctccgtaaaagtacctggtgctgtgtgtttaccattgcaact 2661
Db 61 actttgcccaagtccttctatagtcaaaagtacctggttcggtgctggttagcataaccat 120
Qy 2662 tctggagcttcagacaggttggagctcagggtgcctcttgatatattcgaaggaacacctc 2721
Db 121 tcgggcaattccgacaaactgaaacttcggggcggttcgacgtttctaaaaaagaatttc 180
Qy 2722 ggtaggaaactgaagaatagtcgttcgcgtctctctctctctctctctctctctctctct 2781
Db 181 tccaggaggtttacgttcgagtcgtttgcgctattttcttagggcgtattgtggagatacg 240
Qy 2782 attaaagtcataaggcaatgaagacagaagaatggaagaaacccctgccaatctactgaagat 2841
Db 241 atcaagggttatgaaggcctgaaatcagaggtggtgaaacccactccctcctcctcagcgagat 300
Qy 2842 tctgtatatcgcttcaataatgggggaacgctttctaaactccactgtacgagggcaggtctt 2901
Db 301 tccgtgtacgcgtttcatgacagcgcaatattcgaacgcttctcgtacgtagggcgtgttg 360
Qy 2902 cttggcggttcgaaagcgcgtggttcgagtggttctctctctctctctctctctctctctct 2961
Db 361 ctccgggggtccaaagcttgcgcggtctctctctctctctctctctctctctctctctctctct 420
Qy 2962 gctgcgacgaaggccttttctggcattacgtctgttcttttccacaggttccactattctac 3021
Db 421 actgggaacaaactcttttcagggtctcacatctcttttcccgccggtggtgtttctac 480
Qy 3022 gaccgcggtttaactgaagatgaaaggcttgatgctctggtgcgcacacagaagatgctata 3081
Db 481 gatgaagccttgacgcgccggagagagaggttgatgctacacgcgcgtgacatgctgtg 540
```


CC grapes are grown. Although not lethal, it causes yield losses and
CC reduction in sugar content. The virus encodes several proteins,
CC which may serve as targets for the inhibition of viral infection or
CC replication. These proteins include the 242-248 kD polyprotein (Y58148),
CC encoded by open reading frame (ORF) 1a, and which comprises a proteinase
CC domain (Y58143), a methyltransferase domain (Y58144) and a helicase
CC domain (Y58145); a RNA-dependent RNA polymerase (Y58146), encoded by ORF
CC 1b; and a protein of unspecified function (Y58147), encoded by ORF 11.
CC Nucleotides encoding these proteins, or fragments thereof, and the 5',
CC and 3' untranslated regions (UTRs) of the genome (Z49201-Z49202) are
CC useful for the generation of transgenic plants and plant components.
CC Such transgenic plants may be resistant to viral disease, for example,
CC this property being conferred on the plants via the use of nucleotides
CC encoding truncated or internally deleted proteins, or via the use of
CC antisense nucleotides to inhibit viral gene expression. The nucleotides
CC may additionally be used in the detection of viral nucleic acids in a
CC tissue sample. The proteins can be used to generate antibodies which
CC may be used to detect GLRaV-3 in plant samples. The isolation of
CC GLRaV-3 facilitates the production of agents that reduce the risk of
CC infection or damage by the virus in vineyards.
XX
SQ Sequence 873 BP; 262 A; 163 C; 233 G; 215 T; 0 other;

Query Match 13.0%; Score 873; DB 21; Length 873;
Best Local Similarity 100.0%; Pred. No. 4.6e-273;
Matches 873; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5764 gtggcggaatcgtttaagtccttcgaataataagtcataatgctcccccaggctgcggt 5823
DB 1 gtggcggaatcgtttaagtccttcgaataataagtcataatgctcccccaggctgcggt 60

QY 5824 aaagcagcagcttagtgagcaaatcgttaagtcacccaatagcacagccaccattacg 5883
DB 61 aagacagcagcgttagtgagcaaatcgttaagtcacccaatagcacagccaccattacg 120

QY 5884 gctaattgtgggaattctgagacataaataatggcgtggaagaagagatccgaatttg 5943
DB 121 gctaattgtgggaattctgagacataaataatggcgtggaagaagagatccgaatttg 180

QY 5944 gaagcttcaacagtgctaccacagcttaactccagggtggttaaaactttatgctcaggga 6003
DB 181 gaagcttcaacagtgctaccacagcttaactccagggtggttaaaactttatgctcaggga 240

QY 6004 atgtataaaagggttttggtagagtgagtgacatgatgcatacaggcttactacaacta 6063
DB 241 atgtataaaagggttttggtagagtgagtgacatgatgcatacaggcttactacaacta 300

QY 6064 ggcgtcttcgcaacccggcgcgtcggaaggcctcttttttgagacataaatacagatacca 6123
DB 301 ggcgtcttcgcaacccggcgcgtcggaaggcctcttttttgagacataaatacagatacca 360

QY 6124 ttcatataaagggaagaggtgttttaggaggtattgtctgttttttccaaagaagaa 6183
DB 361 ttcatataaagggaagaggtgttttaggaggtattgtctgttttttccaaagaagaa 420

QY 6184 agcgttgtatacacttcaaaatcgacaggtgtccgttagatgtttgtactgtgttctcc 6243
DB 421 agcgttgtatacacttcaaaatcgacaggtgtccgttagatgtttgtactgtgttctcc 480

QY 6244 tcaatgacgttaagggaaggaaggtgtttaccctgaaaaggctgttagcgttaaggac 6303
DB 481 tcaatgacgttaagggaaggaaggtgtttaccctgaaaaggctgttagcgttaaggac 540

QY 6304 aaaccagtagtaagatcgtgtccaaaggccaaattgaaaccactgatgacgtactaa 6363
DB 541 aaaccagtagtaagatcgtgtccaaaggccaaattgaaaccactgatgacgtactgaa 600

QY 6364 ataaacgctgacgtgtactgttgcattgacccaggttggagagtcggtatgaagaggtcg 6423
DB 601 ataaacgctgacgtgtactgttgcattgacccaggttggagagtcggtatgaagaggtcg 660

QY 6424 ttgaagggaaggaaggaagaaacccagtgatgacagtgatgacagcaggggaaaaaca 6483

DB 661 ttgaagggaaggaaggaagaaacaccagtgatgcagtgcatgaagcacagggaagaaaca 720
QY 6484 ttcaatgagtgaggtattgttttaggagcaagaagccgatgactcccttactactaaacaa 6543
DB 721 ttcaatgagtgaggtattgttttaggagcaagaagccgatgactcccttactactaaacaa 780
QY 6544 ccgcataactctgttggtttgtcgcagacacacacgctcactggtttatgccgctctgagc 6603
DB 781 ccgcataactctgttggtttgtcgcagacacacacgctcactggtttatgccgctctgagc 840
QY 6604 tcaaatgtggacgataaagtcggcgccatattt 6636
DB 841 tcaaatgtggacgataaagtcggcgccatattt 873

RESULT 5
ID Z49204 standard; DNA; 816 BP.
XX
AC Z49204;
XX
DT 07-MAR-2000 (first entry)
XX
XX GLRaV-3 polyprotein methyltransferase domain DNA.
KW Grapevine; leafroll; grapevine leafroll-associated virus 3; GLRaV-3;
KW viral disease; yield loss; sugar content; inhibition; infection;
KW replication; polyprotein; domain; proteinase; methyltransferase;
KW helicase; RNA-dependent; RNA polymerase; untranslated region;
KW transgenic plant; component; resistant; truncation; deletion; antisense;
KW expression; detection; antibody; ds.
XX
OS Grapevine leafroll-associated virus 3.
XX
FH Key Location/Qualifiers
FT mat_peptide 1..816
FT /*tag= a
FT /product= "GLRaV-3 polyprotein methyltransferase domain"
XX
PN W09955880-A1.
XX
XX 04-NOV-1999.
PD
PF 29-APR-1999; 99WO-US09307.
PF
PR 29-APR-1998; 98US-0083404.
PR
XX (CORR) CORNELL RES FOUND INC.
PA
XX Gonsalves D, Ling K;
PI
XX WPI: 2000-062035/05.
DR P-PSDB; Y58144.
DR
XX
PT Newly isolated grapevine leafroll virus protein or polypeptide useful
PT for producing transgenic plants conferring viral disease resistance -
XX
PS Claim 14; Fig 7; 84pp; English.
XX
CC This sequence represents DNA encoding the methyltransferase domain of the
CC polyprotein (Y58148) from the grapevine leafroll-associated virus 3
CC (GLRaV-3). Leafroll is a serious viral disease, occurring wherever
CC grapes are grown. Although not lethal, it causes yield losses and
CC reduction in sugar content. The virus encodes several proteins,
CC which may serve as targets for the inhibition of viral infection or
CC replication. These proteins include the 242-248 kD polyprotein (Y58148),
CC encoded by open reading frame (ORF) 1a, and which comprises a proteinase
CC domain (Y58143), a methyltransferase domain (Y58144) and a helicase
CC domain (Y58145); a RNA-dependent RNA polymerase (Y58146), encoded by ORF
CC 1b; and a protein of unspecified function (Y58147), encoded by ORF 11.
CC Nucleotides encoding these proteins, or fragments thereof, and the 5'
CC and 3' untranslated regions (UTRs) of the genome (Z49201-Z49202) are

CC useful for the generation of transgenic plants and plant components.
CC Such transgenic plants may be resistant to viral disease, for example,
CC this property being conferred on the plants via the use of nucleotides
CC encoding truncated or internally deleted proteins, or via the use of
CC antisense nucleotides to inhibit viral gene expression. The nucleotides
CC may additionally be used in the detection of viral nucleic acids in a
CC tissue sample. The proteins can be used to generate antibodies which
CC may be used to detect GLRaV-3 in plant samples. The isolation of
CC GLRaV-3 facilitates the production of agents that reduce the risk of
CC infection or damage by the virus in vineyards.
XX
SQ Sequence 816 BP; 183 A; 162 C; 234 G; 237 T; 0 other;

Query Match 12.2%; Score 816; DB 21; Length 816;
Best Local Similarity 100.0%; Pred. No. 1.5e-254;
Matches 816; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1378 ctgaagccggaaggaaggaagctgaggaactcttccagagcttcgatacagttc 1437
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 1 ctgaagccggaaggaaggaagctgaggaactcttccagagcttcgatacagttc 60
QY 1438 tccgactcgtcaggagtagtcaccattcgttaatgccatgcggagctgtttcaatgga 1497
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 61 tccgactcgtcaggagtagtcaccattcgttaatgccatgcggagctgtttcaatgga 120
QY 1498 atctttccaggaggtgtgtaattgtgtcttcttcgtatattgggggagcttcacgtat 1557
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 121 atctttccaggaggtgtgtaattgtgtcttcttcgtatattgggggagcttcacgtat 180
QY 1558 catgtcaaaactggccatgtgaactgtcatgtatgtaataatccagcttcagcgttaaatg 1617
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 181 catgtcaaaactggccatgtgaactgtcatgtatgtaataatccagcttcagcgttaaatg 240
QY 1618 gtgaagcggagaaatcaatgagatctcttcttccacagctggggagagattcgtacgtg 1677
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 241 gtgaagcggagaaatcaatgagatctcttcttccacagctggggagagattcgtacgtg 300
QY 1678 tccagtgaccttcaactgaagcggcttcaaaagtctgttcttactgtagtcgagaatcg 1737
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 301 tccagtgaccttcaactgaagcggcttcaaaagtctgttcttactgtagtcgagaatcg 360
QY 1738 cagaactcgatctagagccgatcggttttatgggtgatgtgtacgatatatccccg 1797
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 361 cagaactcgatctagagccgatcggttttatgggtgatgtgtacgatatatccccg 420
QY 1798 cagcaggtagcagaggtctgtgataagaagggtgcgctgtgtttcgacatagctcttatg 1857
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 421 cagcaggtagcagaggtctgtgataagaagggtgcgctgtgtttcgacatagctcttatg 480
QY 1858 ttccccgtgaggtgtgtgtacggttaacggtgaagtttacttggaaagaaactcgatcgttg 1917
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 481 ttccccgtgaggtgtgtgtacggttaacggtgaagtttacttggaaagaaactcgatcgttg 540
QY 1918 gtgaagaggaaggtgattacctggcctacaagtgttggtgaaggtgtggtgagatgtatgaa 1977
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 541 gtgaagaggaaggtgattacctggcctacaagtgttggtgaaggtgtggtgagatgtatgaa 600
QY 1978 cattctctcttaacgtaagcgggtttttcacctttcttcttatgtacgcaccttcgtccggg 2037
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 601 cattctctcttaacgtaagcgggtttttcacctttcttcttatgtacgcaccttcgtccggg 660
QY 2038 aacgtgtttaaagttagatgataaggataaccgttgtgtgtgttaccatcatctcactatgtgt 2097
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 661 aacgtgtttaaagttagatgataaggataaccgttgtgtgtgttaccatcatctcactatgtgt 720
QY 2098 agggctcagaagtcactcgaactgaggttaccgtataggtcggttggtcccgctggttcgtg 2157
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 721 agggctcagaagtcactcgaactgaggttaccgtataggtcggttggtcccgctggttcgtg 780
QY 2158 ggcataatcgctgtgtttcacctgtgttagctggt 2193
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 781 ggcataatcgctgtgtgtttcacctgtgttagctggt 816

RESULT 6

249203
ID 249203 standard; DNA; 360 BP.
XX
AC 249203;
XX
DT 07-MAR-2000 (first entry)
XX
XX GLRaV-3 polyprotein proteinase domain DNA.
XX
XX Grapevine; leafroll; grapevine leafroll-associated virus 3; GLRaV-3;
KW viral disease; yield loss; sugar content; inhibition; infection;
KW replication; polyprotein; domain; proteinase; methyltransferase;
KW helicase; RNA-dependent; RNA polymerase; untranslated region;
KW transgenic plant; component; resistant; truncation; deletion; antisense;
KW expression; detection; antibody; ds.
XX
OS Grapevine leafroll-associated virus 3.
XX
FH Location/Qualifiers
FT mat_peptide 1..360
FT /tag= a
FT /product= "GLRaV-3 polyprotein proteinase domain"
XX
XX W0955880-A1.
XX
XX 04-NOV-1999.
XX
XX 29-APR-1999; 99WO-US09307.
XX
XX 29-APR-1998; 98US-0083404.
XX
XX (CORR) CORNELL RES FOUND INC.
XX
XX Gonsalves D, Ling K;
XX
XX WPI; 2000-062035/05.
DR P-PSDB; Y58143.
XX
PT Newly isolated grapevine leafroll virus protein or polypeptide useful
for producing transgenic plants conferring viral disease resistance -
XX
PS Claim 12; Fig 5; 84pp; English.
XX
CC This sequence represents DNA encoding the proteinase domain of the
CC polyprotein (Y58148) from the grapevine leafroll-associated virus 3
CC (GLRaV-3). Leafroll is a serious viral disease, occurring wherever
CC grapes are grown. Although not lethal, it causes yield losses and
CC reduction in sugar content. The virus encodes several proteins,
CC which may serve as targets for the inhibition of viral infection or
CC replication. These proteins include the 242-248 kD polyprotein (Y58148),
CC encoded by open reading frame (ORF) 1a, and which comprises a proteinase
CC domain (Y58143), a methyltransferase domain (Y58144) and a helicase
CC domain (Y58145); a RNA-dependent RNA polymerase (Y58146), encoded by ORF
CC 1b; and a protein of unspecified function (Y58147), encoded by ORF 11.
CC Nucleotides encoding these proteins, or fragments thereof, and the 5'
CC and 3' untranslated regions (UTRs) of the genome (249201-249202) are
CC useful for the generation of transgenic plants and plant components.
CC Such transgenic plants may be resistant to viral disease, for example,
CC this property being conferred on the plants via the use of nucleotides
CC encoding truncated or internally deleted proteins, or via the use of
CC antisense nucleotides to inhibit viral gene expression. The nucleotides
CC may additionally be used in the detection of viral nucleic acids in a
CC tissue sample. The proteins can be used to generate antibodies which
CC may be used to detect GLRaV-3 in plant samples. The isolation of
CC GLRaV-3 facilitates the production of agents that reduce the risk of
CC infection or damage by the virus in vineyards.
XX
SQ Sequence 360 BP; 90 A; 97 C; 100 G; 73 T; 0 other;

Query Match		5.4%;	Score 360;	DB 21;	Length 360;
Best Local Similarity		100.0%;	Pred. No. 1.7e-106;		
Matches 360;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	253	gtcagcggtcagtcagcggtgagaggggagtgtaagaaggtcttgatgagggcaagg	312		
Db	1	gtcagcggtcagtcagcggtgagaggggagtgtaagaaggtcttgatgagggcaagg	60		
QY	313	acctcaactccgaactccagctgctgatttcgacgtcgtattcgaagctgtttct	372		
Db	61	acctcaactccgaactccagctgctgatttcgacgtcgtattcgaagctgtttct	120		
QY	373	aatgcattactgtctacactaccaccgggtagtcggctatgccccctcaagcgcgag	432		
Db	121	aatgcattactgtctacactaccaccgggtagtcggctatgccccctcaagcgcgag	180		
QY	433	cagcctaaaccggctgttaagaagaatgagcagagaagcccaaacggcgaacgtcaacattgg	492		
Db	181	cagcctaaaccggctgttaagaagaatgagcagagaagcccaaacggcgaacgtcaacattgg	240		
QY	493	gcigttaagcccaacagctgttgcgctccagctaccacttcttaaaaaacaggaagcactg	552		
Db	241	gcigttaagcccaacagctgttgcgctccagctaccacttcttaaaaaacaggaagcactg	300		
QY	553	gagccagcgcaatcagtcaccacaacagtcgttgaggagaagcgcccttgacgtttggc	612		
Db	301	gagccagcgcaatcagtcaccacaacagtcgttgaggagaagcgcccttgacgtttggc	360		
RESULT 7					
ID	A08690	A08690 standard; DNA; 931 BP.			
AC	XX	A08690;			
XX	DT	19-JUL-2000 (first entry)			
XX	DE	PMwAV-2 ORF 1a DNA encoding a helicase.			
XX	DE				
KW	XX	Pineapple mealybug wilt virus 2; PMwAV-2; ORF 1a; helicase;			
KX	XX	transgenic pineapple; resistance; antiviral; ss.			
XX	OS	Pineapple mealybug wilt virus 2.			
XX	XX				
FT	Key	Location/Qualifiers			
FT	mat_peptide	2..928			
FT	FT	/*tag- a			
XX	XX	WO200017372-A2.			
XX	XX				
XX	XX	30-MAR-2000.			
XX	XX				
XX	XX	22-SEP-1999; 99WO-US22152.			
XX	XX				
XX	XX	23-SEP-1998; 98US-0101461.			
XX	XX				
XX	XX	(UYHA-) UNIV HAWAII.			
PA	PA	(UYFL) UNIV FLORIDA.			
XX	XX				
PI	PI	Hu JS, Karasev AV, Dawson WO, Melzer M;			
XX	XX				
XX	XX	WPI; 2000-283596/24.			
DR	DR	P-PSDB; Y91969.			
XX	XX				
XX	XX				
PT	PT	Isolated pineapple mealybug wilt virus proteins and polypeptides,			
XX	XX	useful for protecting pineapples against the virus			
XX	XX				
PS	PS	Claim 57; Page 96-97; 112pp; English.			
XX	XX				
XX	XX	Pineapple mealybug wilt virus 2 (PMwAV-2), open reading frame (ORF) 1a			
CC	CC	is incomplete but encodes a protein which encompasses all motifs			
XX	XX	characteristic of viral helicases. The DNA and protein sequences			
CC	CC	are useful for production of transgenic pineapple plant cultivars and			

CC		also in transformation methods to impart resistance against the virus to				
CC		pineapple plants. Mealybug wilt is a major problem limiting profitable				
CC		pineapple production in many pineapple growing areas worldwide, the				
CC		present invention aims to overcome this problem.				
XX	SQ	Sequence 931 BP; 262 A; 173 C; 242 G; 254 T; 0 other;				
Query Match		2.4%;	Score 162.8;	DB 21;	Length 931;	
Best Local Similarity		50.9%;	Pred. No. 4.8e-42;			
Matches 386;		Conservative 0;	Mismatches 372;	Indels 0;	Gaps 0;	
Qy	5957	gtgtaccacagcttaactccaggtgtgtaaactttatctgcaggggaatgtataaaagg	6016			
Db	174	gtgctcgtagcttaactcacaattatgaactgcgaaggagtaataaattataccctgcg	233			
Qy	6017	ttttggtgagtagagtgacatgcatcaagcttactacaactaggcgtcttcgcaa	6076			
Db	234	ccctagttgacgaatgtatttgatgcacaagggtctgtgtgtgtgggggtattcagtt	293			
Qy	6077	ccggcgcgtcggaagcctctttttggagacataaatcagataccattcataaacagg	6136			
Db	294	ctggagcaagaagagctatactcagggagataccaaccagatccctctcaaatagag	353			
Qy	6137	agaagtggtttaggatggtgtgctgtttttgttccaaagaaggaacgctgtgtataca	6196			
Db	354	agaagtggttttactctaaggagggtgtgactgtccaggtaaagatgaattatttaca	413			
Qy	6197	cttctaaatcgctacaggtgtccgttagatgtttgtctacttgttgtctcaatgaccgtaa	6256			
Db	414	catcagagctctacagatgtcctgcgaggtttgtgtgtgtggtcaagctcactcaaggcgc	473			
Qy	6257	ggggaacggaagtggttacctgaaaaagtcgttagcgtgaaggacaaacacagtagtaa	6316			
Db	474	aagctgggtctatacgcctacctgaagggtgtgtcatgcacacagcgtgaagtggtgttac	533			
Qy	6317	gatcgtgtccaaaaggccaattggaaacctgatgacgtagctgaaataaaacgctgacg	6376			
Db	534	gcagtttataccaaacgcccgtagtgtacgcggaacaagtatacaattggaagctgacg	593			
Qy	6377	tgtacttgtcatgaccacagcttgagaagtcgggatatgaagggtcgttgtgaagggaanaag	6436			
Db	594	cttatataacattcaagcaggagtgtaaaagaaaaagtcgtgagggcgctacgagctgtag	653			
Qy	6437	gaaaagaacacccagtgatgacagtgcatgaagcacagggaaaaaacattcagtgatgtg	6496			
Db	654	gcaggagggataaagtgtttacaagccatgaggcgcaaggtatgacttttgggcgggtcg	713			
Qy	6497	tattgtttaggagcgaagaaagccgatgactccctatttcaactaaacacgcgcataacttg	6556			
Db	714	tgttatgtagattaaagtgcacactgacgattccgtttttctctgagcctcacattttag	773			
Qy	6557	tgtgtttgtcagacacacacgcctcactggttttatgcgctcttgagctcaaatgttgacg	6616			
Db	774	ttagcactctccagacatacaaatcctgtgtctatgcccactctctagtagtaagtagccgc	833			
Qy	6617	ataagtlcgcgcacatatattagcgcgcgtcaacctcaatcagatccgacgcttgccttc	6676			
Db	834	acaagtgaggtgcggccatagactcagttacgcgtgaaggaggtaaagtacgcgtactta	893			
Qy	6677	acaggttcgccccgcgtggtgtcttctcagggtataga	6714			
Db	894	agaccttgtggcgtcggcgttattctcagcgtgattga	931			
RESULT 8						
ID	A08698	A08698 standard; DNA; 10000 BP.				
XX	AC	A08698;				
XX	DT	19-JUL-2000 (first entry)				
XX	XX					

DE Pineapple mealybug wilt virus genome 2.
XX
KW Pineapple mealybug wilt virus 2; PMWV-2; transgenic pineapple;
KW resistance; antiviral; ss.
XX

OS Pineapple mealybug wilt virus 2.

XX Key Location/Qualifiers

FT CDS 1..931

FT /tag= a

FT /note= "ORF 1a"

FT CDS 879..2558

FT /tag= b

FT /note= "ORF 1b"

FT CDS 3173..3326

FT /tag= c

FT /note= "ORF 2"

FT CDS 3340..4965

FT /tag= d

FT /note= "ORF 3"

FT CDS 5203..6414

FT /tag= e

FT /note= "ORF 4"

FT CDS 7430..8905

FT /tag= f

FT /note= "ORF 5"

FT CDS 8490..7398

FT /tag= g

FT /note= "ORF 6"

FT CDS 8895..9410

FT /tag= h

FT /note= "ORF 7"

FT CDS 9407..9991

FT /tag= i

FT /note= "ORF 8"

XX WO200017372-A2.

XX 30-MAR-2000.

XX 22-SEP-1999; 99WO-US22152.

XX 23-SEP-1998; 98US-0101461.

XX (UYHA-) UNIV HAWAII.

XX (UYFL) UNIV FLORIDA.

XX Hu JS, Karasev AV, Dawson WO, Melzer M;

XX WPI: 2000-283596/24.

XX P-PSDB; Y91975.

XX Isolated pineapple mealybug wilt virus proteins and polypeptides,

XX useful for protecting pineapples against the virus

XX Claim 47; Page 75-78; 112pp; English.

XX This is the Pineapple mealybug wilt virus 2 (PMWV-2) genome which spans
XX 8 open reading frames which may encode four protein products. The DNA
XX and protein sequences are useful for production of transgenic pineapple
XX plant cultivars and also in transformation methods to impart resistance
XX against the virus to pineapple plants. Mealybug wilt is a major problem
XX limiting profitable pineapple production in many pineapple growing areas
XX worldwide, the present invention aims to overcome this problem.

XX Sequence 10000 BP; 2914 A; 1734 C; 2441 G; 2911 T; 0 other;

XX Query Match 2.4%; Score 162.8; DB 21; Length 10000;

XX Best Local Similarity 50.9%; Pred. NO. 2.8e-41;

XX Matches 386; Conservative 0; Mismatches 372; Indels 0; Gaps 0;

XX QY 5957 gtgctaccacagtaactccagggtgtgtaaacctttatctgcagggaagtataaaagg 6016

Db 174 gtctcgtaacagtaactcaaaattatgaactgcaaggagtaagtaattacacgtcg 233
QY 6017 ttttggtgaggtgacatgatgcatacgaagcttactacacactagcgcttctcgcaa 6076
Db 234 cctagtgaacgaatgtattgatgcacaagggtctgttgatgtgggggttccagtt 293
QY 6077 cggcgctcggaagcctctttttggagacataaatcagatcaccattcataaacagg 6136
Db 294 ctggagcaagaagagctataattctcagagagatacaccagatcccttataaatag 353
QY 6137 agaaggttttagatggattgtctgttttttccaaagaaggaaagcgtgtatataca 6196
Db 354 agaagtgttttacttaaggagggtgtgtactgtccaggtlaaagatgaattattaca 413
QY 6197 cttctaaatcgtacaggtgtccgttttagatgtttgtctactgttctctcactaacgcgttaa 6256
Db 414 catcagagcttaccagatgtccgtccgaggtttgtatgtgttcaagctcactcaaggcgc 473
QY 6257 ggggaacgaaagtgttaccctgaaaagtcgttagcggtgaaggacaaaccagtagtaa 6316
Db 474 aagctgggtctaatcgtctacactgaagggtgtgtcatgcaaccagcgtgaagtgtgttac 533
QY 6317 gatcgtgtccaaaggccaatttgaaccactgatgcgtagctgaaataaaacgctgacg 6376
Db 534 gcagtttatccaaacgcccggtagtgtacgcggaacaagtgtacaaattggaagctgacg 593
QY 6377 tgtactgtgcatgaccaggttggaagtcggtatgaagaggtgtgttgaagggaagaag 6436
Db 594 cttatatacattcaagcaggaggtgtaaaagaaaaagtcgtgaggcgctacgagctgtag 653
QY 6437 gaaaagaacaccagtgatgacagtgcatgaagcacaggggaaacacattcagtgatg 6496
Db 654 gcaggaggataaagtgtttacaagcctagagcgcaaggtatgacttttggcggtgcg 713
QY 6497 tattgtttaggacgaagacgagtcgactccctattcactaaacacgcgcataacttg 6556
Db 714 tgtatgtagattaaagtgcactgacgattccgtttttcttctgagcctcacatttag 773
QY 6557 ttggtttgtcgagacacacagctcactgtttatgcgctctgtgagctcaaaagttgagcg 6616
Db 774 ttgactctccagacatacacaaatcctgtgtctatgccactcttagtaagtttagccg 833
QY 6617 ataagtcggacataatattagcgacgctcactcaatcagatcgcgcgctttgtcttc 6676
Db 834 acaagtggtgcgcacatagactcagtcggtacggtgaaggaggtgaagtgcacggtactta 893
QY 6677 acagttcccccgcgctgtgttcttcgaggtatatga 6714
Db 894 agaccttgcgcgcggtgtatttcttcgagctgattga 931

RESULT 9

V99284

ID V99284 standard; cDNA; 8743 BP.

XX V99284;

AC V99284;

XX 26-APR-1999 (first entry)

XX Rupestris stem pitting associated virus RSPV-1 genome.

XX Rupestris stem pitting associated virus RSPV-1 genome.

XX RSPV-1; grape; transgenic plant; disease resistance; ss.

XX Rupestris stem pitting associated virus.

XX Rupestris stem pitting associated virus.

XX Key Location/Qualifiers

FT CDS 62..6546

FT /tag= a

FT /product= replicase

FT /transl_except= (pos:5852..5853, aa:Asp)

FT /note= "this codon has an apparent 1 nucleotide

FT deletion which alters the reading frame;

FT FT this region is specifically claimed in
FT FT Claim 22"
FT CDS 6577..7242
FT /tag- b
FT /product- triple gene block first protein
FT /note- "this region is specifically claimed in
FT Claim 28"
FT CDS 7244..7597
FT /tag- c
FT /product- triple gene block second protein
FT /note- "this region is specifically claimed in
FT Claim 30"
FT CDS 7519..7760
FT /tag- d
FT /product- triple gene block third protein
FT /note- "this region is specifically claimed in
FT Claim 32"
FT CDS 7770..8549
FT /tag- e
FT /product- coat protein
FT /note- "this region is specifically claimed in
FT Claim 25"

XX WO9852964-A1.
XX 26-NOV-1998.
XX 20-MAY-1998; 98WO-US10391.
XX 17-DEC-1997; 97US-0069902.
XX 20-MAY-1997; 97US-0047147.
XX (CORR) CORNELL RES FOUND INC.
XX Consalves D, Meng B;
XX WPI; 1999-045297/04.
XX P-PSDB; W87724-28.
XX Isolated proteins from Rupestris stem pitting-associated virus and
XX related nucleic acid - vectors, host cells and transgenic Vitis
XX cultivars that are resistant to the virus
XX Claim 63; Page 9-14; 163pp; English.
XX This is the nucleotide sequence of the Rupestris stem pitting
XX associated virus RSPav-1 genome. It is derived from cDNA clones
XX of a dsRNA that was associated with Rupestris stem pitting. The
XX sequence includes 5 open reading frames (ORF1-5) that encode the
XX viral replicase, the 3 proteins of the RSPav-1 triple gene block
XX and the coat protein (see W87724-28). Substantial portions of the
XX genomes of 2 other RSPav strains, RSP47-4 and RSP158, have also
XX been sequenced (see V99285-86), as have portions of 7 other RSPav
XX strains (see V99287-93). Also claimed are: an expression system
XX comprising nucleic acids encoding RSPav proteins or polypeptides;
XX transformed host cells; transgenic grape (Vitis) scion or rootstock
XX cultivars that comprise DNA encoding a replicase, coat protein or
XX triple gene block protein of RSPav; a method of imparting RSPav-1
XX resistance to Vitis cultivars by transformation using such DNA;
XX and primers and probes (see V99294-305) for detection of RSPav
XX sequences.
XX Sequence 8743 BP; 2430 A; 1677 C; 2076 G; 2560 T; 0 other;

Query Match 0.8%; Score 52.8; DB 20; Length 8743;
Best Local Similarity 50.7%; Pred. No. 1.5e-05;
Matches 154; Conservative 0; Mismatches 147; Indels 3; Gaps 1;

QY 4636 accagttctcaaacgattaccctttgtcgataacttgaagaggaggtgcggtcttt 4695
DB 2234 auctgggtgcgaagaacattgttttaagacagcgtgagaggcgggtgcatactc 2293

QY 4696 ttctcaaaagtggp---tgagggggtatacctataatgggtgtagccatgtttcaccagg 4752
DB 2294 ttctcaaaaccaggaaatttcattcatgtcattacaatgggtggagccacacaagcttaggg 2353
QY 4753 tggctctgtgcccctagaggatatttaacggcaattaaagtaccacaagcgtcttcgaccac 4812
DB 2354 tggccaaagtctcattggtcagattcttaagctccactggtggcgttaattactacaattca 2413
QY 4813 tgttttagtcagaagtacaaagtgggtggaggcgtaccattccacgctgtagcagaggag 4872
DB 2414 tgcttggtcagatctatgaggaaaaattcaaatggctcttcataagagatgatgagagt 2473
QY 4873 tgctatccatcagataaacctattcttgacggtcgaactctgtgggaaggcaaaactctcg 4932
DB 2474 tctatgaattgggcacaaagtgttgactgttaatttaatcggtcagcaactttcact 2533
QY 4933 acta 4936
DB 2534 atta 2537
RESULT 10
A10595
ID A10595 standard; DNA; 6741 BP.
XX AC A10595;
XX DT 29-JUN-2000 (first entry)
XX DE Gene encoding a subunit of cellulose synthase.
XX KW Cellulose synthase; cellulose production; increase yield; ds.
XX OS Vigna angularis.
XX PN JP2000060568-A.
XX PD 29-FEB-2000.
XX PF 26-AUG-1998; 98JP-0239998.
XX PR 26-AUG-1998; 98JP-0239998.
XX PA (MIZU/) MIZUNO K.
XX PA (OJIP) OJI PAPER CO.
XX WPI; 2000-342371/30.
XX P-PSDB; Y85180.
XX A gene encoding a cellulose synthetic equipment - for the improvement
XX in the amount of cellulose synthesised in a plant body
XX Claim 2; Page 22-31; 32pp; Japanese.
XX This sequence represents a gene encoding a subunit of the cellulose
XX synthase complex of Vigna angularis. The invention relates to subunits of
XX cellulose synthetic equipment, that can be used to increase the amount of
XX cellulose synthesised by a plant. The proteins and genes encoding them
XX can also be used to improve the properties of the cellulose being
XX produced by a plant.
XX Sequence 6741 BP; 1712 A; 870 C; 1468 G; 1374 T; 1317 other;

Query Match 0.6%; Score 41; DB 21; Length 6741;
Best Local Similarity 19.1%; Pred. No. 0.086;
Matches 93; Conservative 148; Mismatches 247; Indels 0; Gaps 0;

QY 4079 acgctatgctttgttgaaagtgaataagggtgcacgatgtcttcttccacaagt 4138
DB 6024 asgyysrscysydhvhaashngasnaasasaaaaaagaavagaasngyysyasha 6083
QY 4139 ctagtgggaaacggtcgtgaggttgaggtgagcgcaagggtgtgtccccaagaagcg 4198


```

Db 6084 sygstrtyrvagyyaagysysrgarggasgysgnarghgnsmtysgaaaasys 6143
Qy 4199 tcgtcgtaggcgcgcacacaagaagggaagagctgcagatggttaacacagcacaaa 4258
Db 6144 tyrgngyaaasntyrvaasasasasasasasysgshsrhgythrtthrsrctysv 6203
Qy 4259 ccgcggtcaacgaagcgacagagccagtcacagtcacgtcttgcagttccccacagg 4318
Db 6204 amtargasrasngyvasrargysrgyghvaahsrthrrggaasrargaaasrgmtasngy 6263
Qy 4319 ctgatattccaaaggtccaccagtcacaggtacatgctcgaagaagaagtgaacaagaag 4378
Db 6264 ysmtvaasrystyrivathraagnargysgasargaaargnaaghaagmntarg 6323
Qy 4379 taccatggcgactgttgcggcgccacgcacatctcgtatgagaacccgcgcccaagt 4438
Db 6324 rvagymtrrsrvagytargvarmttyrrgyrgyngnhtyrgyngyrraarsrg 6383
Qy 4439 ttacgactcgtgtgtaagataattgacaagggcgaagccgcgtcctcatgtgctgaga 4498
Db 6384 naagyhgtyrgngngnrgymtargyaaarvarasnhhvarmtvagnngyngn 6443
Qy 4499 aaaaacaggtacaagtcgacagcccaacacagagaggtttgcagatcaatgaagcaagg 4558
Db 6444 gygnargygygargargavaagngnsrgnrvmtmcrngnmtrarggyargvaty 6503
Qy 4559 ccggtaaa 4566
Db 6504 raigttyr 6511

```

RESULT 11

Q12707/c

Q12707 standard; DNA; 5510 BP.

Q12707;

27-SEP-1991 (first entry)

Rice actin gene RAC1.

Promoter; transgenic plant; monocotyledonous; ss.

Oryza sativa variety IR26.

Key Location/Qualifiers

promoter 1..2071

repeat_region 1027..1038

repeat_region 1078..1088

repeat_region 1301..1333

repeat_region 1352..1386

repeat_region 1465..1505

repeat_region 1609..1617

repeat_region 1647..1653

repeat_region 1650..1728

repeat_region 1650..1728

repeat_region 1650..1728

repeat_region 1650..1728

repeat_region 1650..1728

repeat_region 1650..1728

repeat_region 1650..1728

repeat_region 1650..1728

repeat_region 1650..1728

repeat_region 1650..1728

repeat_region 1650..1728

repeat_region 1650..1728

repeat_region 1650..1728

repeat_region 1650..1728

repeat_region 1650..1728

repeat_region 1650..1728

repeat_region 1650..1728

repeat_region 1650..1728

repeat_region 1650..1728

repeat_region 1650..1728

repeat_region 1650..1728

repeat_region 1650..1728

repeat_region 1650..1728

repeat_region 1650..1728

repeat_region 1650..1728

```

FT repeat_region 1650..1676
FT /tag= j
FT /rpt_type= tandem
FT /note= "GC-rich - repeated (A/T)CC triplets"
FT misc_RNA 1728..1731
FT /tag= k
FT /label= 5' acceptor splice site
FT intron 1729..2043
FT /tag= l
FT misc_RNA 2008..2012
FT /tag= m
FT /label= putative branch point splice site
FT misc_RNA 2034..2044
FT /tag= n
FT /label= 3' donor splice site
FT CDS 2051..3600
FT /tag= o
FT misc_RNA 1650..3741
FT /tag= p
FT /label= genomic clone pRac1

```

W09109948-A.

11-JUL-1991.

04-JAN-1991; 91WO-US00073.

05-JAN-1990; 90US-0461490.

(CORR) CORNELL RES FOUNDATION.

Wu R, Mcelroy D;

WPI: 1991-222901/30.

Rice actin gene promoter - used to confer high level expression

of foreign genes in transgenic rice and other plants

Claim 8; Page 10; 47pp; English.

The sequence was obt'd. from a clone isolated from a rice genomic

library screened with a heterologous actin probe. The CDS encodes

a 377 amino acid actin protein of 41.9 kD. The gene is interrupted

by three introns (position not disclosed). Analysis also revealed

an additional intron 5' of the translation initiation region.

This 5' intron separates a 79 bp GC-rich 5' noncoding exon from

the exon coding the translation initiation codon. This is one of

few cases of a 5'-noncoding exon in a plant gene. The promoter

region can be used to confer high level expression of foreign

genes in transgenic rice and other agronomically important plants.

It is five times more active than the maize Adh1 promoter in trans-

formed rice.

Sequence 5510 BP; 1466 A; 1262 C; 1163 G; 1619 T; 0 other;

Query Match

Best Local Similarity 0.6%; Score 37.8; DB 12; Length 5510;

Matches 69; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Qy 4667 ataacttgaaggagagagtgoggtttttctcaagctgggtgaggggtatacctata 4726

Db 4174 ACAACTTTTAAAGCGCAAGCGTCTCTGTGCTGAAAGCTGACGATGAGAATATCTG 4115

Qy 4727 atggtgttagcctatttcacaggtgcctcgtccctagaggatatttcaacggcaa 4786

Db 4114 ATCGTGTGATGTCATCTTTCATCATCTGTCGCAACAATAGCCCAAGAAAGATGACAGAAG 4055

Qy 4787 t 4787

Db 4054 T 4054


```
A08684
ID A08684 standard; DNA; 1799 BP.
XX
AC A08684;
XX
DT 19-JUL-2000 (first entry)
XX
DE PMWAV-1 helicase DNA.
XX
KW Pineapple mealybug wilt virus 1; PMWAV-1; ORF 1a; helicase;
KW transgenic pineapple; resistance; antiviral; ss.
XX
OS Pineapple mealybug wilt virus 1.
XX
FH Key Location/Qualifiers
FT mat_peptide 3..1799
FT /tag= a
FT /product= helicase
XX
PN WO200017372-A2.
XX
PD 30-MAR-2000.
XX
PF 22-SEP-1999; 99WO-US22152.
XX
PR 23-SEP-1998; 98US-0101461.
XX
PA (UYHA-) UNIV HAWAII.
PA (UYFL) UNIV FLORIDA.
XX
PI Hu JS, Karasev AV, Dawson WO, Melzer M;
XX WPI; 2000-283596/24.
DR P-PSDB; Y91963.
XX
PT Isolated pineapple mealybug wilt virus proteins and polypeptides,
PT useful for protecting pineapples against the virus
XX
PS Claim 38; Page 79; 112pp; English.
XX
CC Pineapple mealybug wilt virus 1 (PMWAV-1), open reading frame (ORF) 1a
CC encodes a protein with all 8 motifs conserved in the so-called viral
CC helicases. The DNA sequences and related proteins are useful for
CC production of transgenic pineapple plant cultivars and also in
CC transformation methods to impart resistance against the virus to
CC pineapple plants. Mealybug wilt is a major problem limiting profitable
CC pineapple production in many pineapple growing areas worldwide, the
CC present invention aims to overcome this problem.
XX
SQ Sequence 1799 BP; 565 A; 323 C; 455 G; 456 T; 0 other;

Query Match 0.5%; Score 36.2; DB 21; Length 1799;
Best Local Similarity 45.8%; Pred. No. 1.2;
Matches 125; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

QY 6002 gaatgtataaaagggtttgttgatgagtgacatgatgcataaggttactacaac 6061
Db 1078 gactttataaaacgttgctgatagacgaatgctacatgacacgcggcatattgcgag 1137
QY 6062 taggcgtcttcgcaaccgcgcgctggaagccctctttttggagacataaatcaatcac 6121
Db 1138 gtataatccgcgtgtgaagccggaagagtgtgtactacgtgtgtagcgacaaagtgc 1197
QY 6122 catctataaacagggagaggtgttaggtgattgctgtctgtttgttccaaagaag 6181
Db 1198 cctttatacaacgaatacaattgctaataatgacaaatacttttgaaccgagtcctcg 1257
QY 6182 aaagcgtgtgtacacacctcttaactgtacaggtgctcgttagatgtttgtctactttgt 6241
Db 1258 gcaattattcagagatgttgataaaccagcggtgtcccgacagacatctgttggcgatgt 1317
QY 6242 cctcaatgaccgtgaagggaacggaagaagtgtt 6274

Db 1318 ccaatgttaacaatgggaagaaaggagatcgtt 1350

RESULT 15
A08697
ID A08697 standard; DNA; 5217 BP.
XX
AC A08697;
XX
DT 19-JUL-2000 (first entry)
XX
DE Pineapple mealybug wilt virus genome 1.
XX
KW Pineapple mealybug wilt virus 1; PMWAV-1; transgenic pineapple;
KW resistance; antiviral; ss.
XX
OS Pineapple mealybug wilt virus 1.
XX
FH Key Location/Qualifiers
FT prim_transcript 1..1799
FT /tag= a
FT /note= "contains ORF 1a"
FT prim_transcript 1783..3350
FT /tag= b
FT /note= "contains ORF 1b"
FT prim_transcript 3360..3499
FT /tag= c
FT /note= "contains ORF 2"
FT CDS 3183..5012
FT /tag= d
FT /note= "ORF 3"
XX
PN WO200017372-A2.
XX
PD 30-MAR-2000.
XX
PF 22-SEP-1999; 99WO-US22152.
XX
PR 23-SEP-1998; 98US-0101461.
XX
PA (UYHA-) UNIV HAWAII.
PA (UYFL) UNIV FLORIDA.
XX
PI Hu JS, Karasev AV, Dawson WO, Melzer M;
XX WPI; 2000-283596/24.
DR P-PSDB; Y91963, Y91964, Y91965, Y91966.
XX
PT Isolated pineapple mealybug wilt virus proteins and polypeptides,
PT useful for protecting pineapples against the virus
XX
PS Claim 47; Page 73-75; 112pp; English.
XX
CC This is the Pineapple mealybug wilt virus 1 (PMWAV-1) genome which spans
CC 4 open reading frames which may encode four protein products. The DNA
CC and protein sequences are useful for production of transgenic pineapple
CC plant cultivars and also in transformation methods to impart resistance
CC against the virus to pineapple plants. Mealybug wilt is a major problem
CC limiting profitable pineapple production in many pineapple growing areas
CC worldwide, the present invention aims to overcome this problem.
XX
SQ Sequence 5217 BP; 1565 A; 925 C; 1267 G; 1460 T; 0 other;

Query Match 0.5%; Score 36.2; DB 21; Length 5217;
Best Local Similarity 45.8%; Pred. No. 2.6;
Matches 125; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

QY 6002 gaatgtataaaagggtttgttgatgagtgacatgatgcataaggttactacaac 6061
Db 1078 gactttataaaacgttgctgatagacgaatgctacatgacacgcggcatattgcgag 1137
```

QY 6062 taggcgtcttcgcaaccggcgcgtcgcgaaggccctctcttttttggagacataaaatcagatac 6121
Db 1138 gtataatcgccgctgtgaagccgcgaagagtggtacttttacggtgatagggcgacaagtgc 1197
QY 6122 cattcataaacagaggagaggtgttttaggatggattgctgtgttttttttccaaagaagg 6181
Db 1198 cctttatcacagaatcaaatgctaaatgacacaaatcccttttgaaccgagtcctcg 1257
QY 6182 aaagcgttggtatacacacttctctaaatcgtacaggtgtccgttagatgtttgctacttgtgt 6241
Db 1258 gcaattattcagagatgttgataaccaggcgcgtgcccgacagacatctgttggcggatgt 1317
QY 6242 cctcaatgaccgtaaggggaaacgcggaagagttt 6274
Db 1318 ccaatgttaacaatgggaagaagagatcgtt 1350

Search completed: February 14, 2001, 17:04:45
Job time: 6787 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 14, 2001, 16:54:31 ; Search time 127.18 Seconds
(without alignments)
8507.865 Million cell updates/sec

Title: US-09-301-906-3
Perfect score: 6714
Sequence: 1 atggactacattgcccatt.....gttgctttcagggtatatga 6714

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 280836 seqs, 80580151 residues

Total number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	2557	38.1	4173	2	US-08-770-544-1
2	54.6	0.8	7218	1	US-08-232-463-14
3	52.8	0.8	6485	3	US-09-081-320-2
4	52.8	0.8	8743	3	US-09-081-320-1
5	37.8	0.6	5843	1	US-08-144-6028-4
6	35.4	0.5	210	1	US-08-036-555B-157
7	35.4	0.5	210	1	US-08-469-569-157
8	35.4	0.5	210	1	US-08-249-322A-157
9	35.4	0.5	210	1	US-08-469-526A-157
10	35.4	0.5	210	2	US-08-734-591A-157
11	35.4	0.5	210	2	US-08-469-660-157
12	35.4	0.5	210	3	US-08-341-018-65
13	35.4	0.5	210	3	US-08-470-335-157
14	35.4	0.5	210	4	PCT-US94-05083C-153
15	35.4	0.5	210	4	PCT-US95-06846A-157
16	34.6	0.5	1239	1	US-08-349-696-22
17	34.6	0.5	1239	1	US-08-233-009-22
18	34.6	0.5	1239	1	US-08-560-231-22
19	34.6	0.5	1239	3	US-09-080-704A-22
20	33.2	0.5	2437	3	US-08-456-647B-3
21	33.2	0.5	2437	2	US-08-237-401A-3
22	33	0.5	2389	2	US-08-372-652-8
23	33	0.5	2389	4	PCT-US95-16311-8
24	33	0.5	6744	1	US-08-119-125A-2
25	32.8	0.5	1069	3	US-09-136-442-2
26	32.6	0.5	3323	4	US-07-624-295-14
27	32.4	0.5	774	4	PCT-US91-05766-1
28	32.4	0.5	1167	3	US-08-691-563C-61

29	32.4	0.5	5340	3	US-08-492-459-9	Sequence 9, Appl
30	32.4	0.5	5340	3	US-08-423-752-9	Sequence 9, Appl
31	32.4	0.5	5340	3	US-08-945-994-2	Sequence 2, Appl
32	32	0.5	693	1	US-08-168-091A-3	Sequence 3, Appl
33	32	0.5	2730	1	US-08-344-536-1	Sequence 1, Appl
34	32	0.5	2730	3	US-08-920-562-1	Sequence 1, Appl
35	31.4	0.5	6725	3	US-08-949-386-36	Sequence 36, Appl
36	31.4	0.5	6725	3	US-08-450-562-36	Sequence 36, Appl
37	31.2	0.5	5232	3	US-08-972-927-1	Sequence 1, Appl
38	31.2	0.5	9936	3	US-08-972-927-2	Sequence 2, Appl
39	31	0.5	6519	1	US-08-588-985-1	Sequence 1, Appl
40	31	0.5	6519	1	US-08-971-988-1	Sequence 1, Appl
41	31	0.5	12720	1	US-08-403-866-11	Sequence 11, Appl
42	30.8	0.5	885	1	US-08-411-706-3	Sequence 3, Appl
43	30.8	0.5	2658	2	US-08-773-608A-1	Sequence 1, Appl
44	30.8	0.5	3370	1	US-08-060-822A-5	Sequence 5, Appl
45	30.8	0.5	3370	4	PCT-US94-05257-5	Sequence 5, Appl

ALIGNMENTS

RESULT 1
US-08-770-544-1
; Sequence 1, Application US/08770544
; Patent No. 5907085
; GENERAL INFORMATION:
; APPLICANT: Gonsalves, Dennis
; APPLICANT: Ling, Kai-Shu
; TITLE OF INVENTION: GRAPEVINE LEAFROLL VIRUS PROTEINS AND
; TITLE OF INVENTION: THEIR USES
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/770,544
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60009008
; FILING DATE: 21-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/621
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4173 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-770-544-1

Query Match 38.1%; Score 2557; DB 2; Length 4173;
Best Local Similarity 76.3%; Pred. No. 0;
Matches 3187; Conservative 0; Mismatches 980; Indels 12; Gaps 3;
QY 2542 gtctcgtctatgttaagcaatggtgcacataacttaacgttttttgagacgttatg 2601

QY 4756 cctcgctccctagagatatcttaacggcaattaaagtacccaagcgtcttcgaccactgt 4815
DB 2215 CCTCGTGCCTTAGAGGATATCTTAAAGGCAATTAAGTACCAAGCGTCTTCGACCACCTGT 2274
QY 4816 ttagtcagaaagLacaagaatgggtggagggcggtaccattccacgctgatgacgagaggtgc 4875
DB 2275 TTAGTCAGAAAGTACAAAGATGGGTGGAGGCGTACCATTCCACGCTGATGACGAGGAGTGC 2334
QY 4876 tatecatcagataacccattcttgacggtcaatctcgtgggggaagcgaactctcgcact 4935
DB 2335 TATCCATCAGATAACCCATCTATCTGACGGTCAATCTCGTGGGAAGCAAACTTCTCGACT 2394
QY 4936 aagtgcaggaaggttgtaaggtcattggctataaaacgttaagctcgggtgaactatttctt 4995
DB 2395 AAGTCAGGAAGGAGGTGGTGAAGGTCAATGGTCAATAAAGCTAGCTTCGGGTGACTATTTCTT 2454
QY 4996 atgccttcgggtttcaaaaggaagcagcttgcattcagtaaaactccatcacgaagggcgc 5055
DB 2455 ATGCCTTGGGTTTCAAGGACGCACTTGCATTCAAGTAACCTCCATCGACGAAGGCGC 2514
QY 5056 atcagtttgacgttcggggcaactcgccgctctcttggtagcgaggtatgttgaggtta 5115
DB 2515 ATCAGTTTGACGTTTCAAGGCACTCGCGCGCTCTTGGTGTAGCGAGGATGTTGCAGTTA 2574
QY 5116 gccggcggtgctcgatgagaagtcaccaggtgtttccaaacacagcaacacagagccaa 5175
DB 2575 GCCGGCGGCGTGCAGATGAGAAGTCACCAAGGTGTTCCAAACAGCAACACACAGGCCAA 2634
QY 5176 ggtgctaccagaacaatcacacaaatcggggggcaagcctctatctgagggaaagtgtt 5235
DB 2635 GGTGCTACCAAGACATCACACCAAAATCGGGGGGCAAGCCTCTACTGAGGGAAGTGT 2694
QY 5236 agggaaagtcaggggaggtgcacatactcgatagtggtgcaacaaagattacgttaggaag 5295
DB 2695 AGGGAAGTCAAGGGAGGTGCACATCTCGATATGGTGGCAACAAAGATTACGTTAGGAAG 2754
QY 5296 tggagtgctcaggctgataatcacgtagtcaggtgttgaacctgactacacacccaatg 5355
DB 2755 TGTGAGTGGCTCAGGCGTGATAATCCAGTATGGCTCTTAAACCTGGCTACACCCCAATG 2814
QY 5356 acatttgaagtgttaaaacgggacctctgaagatgcgctgagtgacttgaaagtat 5415
DB 2815 ACATTTGAAGTGTTAAGCCGGGACCTCTGAAGATGCCGCTGAGTACTTTGAAGTAT 2874
QY 5416 ctggctataggcattgagaggacatacagggcgcttcttatggctagaaatattgcgctc 5475
DB 2875 CTGGCTATAGGCATTTGGAGGACATACAGGCGCTTGCTTATGGCTAGAAATATTGCCGTC 2934
QY 5476 actaccgcgaaggtgttctgaaagtacctaatacagtttatgaatcactaccggccttt 5535
DB 2935 ACTACCCGCCGAAGGTGTTCTGAAGTACCTTAATCAAGTTTATGAATCACTTACCGGGCTTT 2994
QY 5536 cagctttacaagtcgggcagagatctcattttcattcaacaagaacagcgcttgagtg 5595
DB 2995 CAGCTTTACAGTTCGGGCACAGATCTCATTTTTCATTCAACAGACGCGCTTCGCTGTG 3054
QY 5596 agagacctaccgtacgtactcatagctgaaaggtatcttttaccgaagggcaagatgtc 5655
DB 3055 AGAGACCTACCGTACGTATTATAGCTGAGAAAGGTATTTTATCAAGGGCAAGATGTC 3114
QY 5656 gacgcggttgagcttttgggcacaatctgtctgtatgcagagataacttggttttccac 5715
DB 3115 GACCGGCTAGTAGCTTTGGCGCAATCTGTCCGTATGTGATGATATATTGGTTTTCAT 3174
QY 5716 gatgcatttaatttgatagtgacatgaagtcgctcgatgcgagcgtatggtggcgaatcg 5775
DB 3175 GATGCTATTAAATTGATGGTGACATGAAGATGCTCGATGTGGTATGTTGGGTGAATCA 3234
QY 5776 tttaagcttcgaataataagtgctataatgctcccccggtggcggtgaagacgacgacg 5835
DB 3235 TTTAAGCTGCTGAATAACAAATGCTATATGCTCCCCAGGTGGCGGTAAAGACGATG 3294

QY 5836 ttagtgagcgaattcgttaagtaccaccaatagcacagccaccattacgggtaattgtgga 5895
DB 3295 CTAGTGGACAAATTTGTCAAGTCAACCAATAGCAGCGCCACCATTAGCGTAAACGGGA 3354
QY 5896 agtctcagagacataaataatggtgaggaagagagatccgaatttggaggtcctcaac 5955
DB 3355 AGTCTGAGGACATAAATATGCGGTGGAAGAGAGATCCGAATTTGGAAAGGTCTCAAC 3414
QY 5956 agtgcaccacagtttaactccagggtgggttaaaactttatcgtcagggggaatgtataaag 6015
DB 3415 AGTGTACCAACAGTTAACTCCAGGTGCTTAACTTTATTCTCAGGGGAATGTATAAAGG 3474
QY 6016 gttttgggtgagtgaggtgacatgacatcaaggcttactacaactagcgcttctcgca 6075
DB 3475 GTTTTGGTGGATGAGGTGTACATGATCATCAAGGCTTACTACAACATAGGCGCTTCCGA 3534
QY 6076 accggcgctcggaagcctctttttggagacataaatacagataccattcataaacagg 6135
DB 3535 ACCGGCGCTCGGAAGCCCTCTTTTGGAGACATAAATCAGATACCATTCATAAACCGG 3594
QY 6136 gagaaggtgttttaggattggtgctgtttttgttccaaagaaggaagcgtttgtatatac 6195
DB 3595 GACAAGGTGTTTAGGATGATTGCTGTATTTGTTCCAAAGGAAGAAACGGTTGTATAC 3654
QY 6196 acttctaaatcgttacagggtgcggttagatggtttgtacttgttgcctcaatgacctga 6255
DB 3655 ACTTCTAAATCATACAGGTGCTCGTTAGATGTTTGTCTACTTGTTCCTCAATGACCGTA 3714
QY 6256 aggggaacggaaggtgttaccctgaaaggtcgttagcggtgaagacaacacagtagta 6315
DB 3715 AGGGGAACGGAAGGTGTACCTGAAAGGTGTTAGCCGTTAGGACAAACCAAGTAGTA 3774
QY 6316 agatcgtctgtccaaagggccaaattggaaacactgatgacgttagctgaaataaaacgctgac 6375
DB 3775 AGATCGCTGTCCAAAGGCCAAATTGGAACCACTGATGACGTAGTGAATAAAGCGCTGAC 3834
QY 6376 gtgtactgtgcatgacccagttggagaagtcggatatgaagagggtcgttgaaggggaaaa 6435
DB 3835 GTGTACTGTGTCATGATGCCAGTTGGAGAGTCGGATATGAAGAGGTGCTTTGAAGGGGAAAA 3894
QY 6436 ggaagaagaacacagtgatgacgtgcatgaagcacagggagaaacattcagtgatgtg 6495
DB 3895 GGAAGAACAACACAGTGTGATGACAGTGCATGAAGCACAGGGGAAAAACATTCAGTGATGTG 3954
QY 6496 gtattgttaggcgaagaagccgatgactccttacttaacaaacccgcataactt 6555
DB 3955 GTATTGTTTAGGACGAAGAACCGGATGACTCCCTATTACTAAACAACCCGATATACTT 4014
QY 6556 gttgggttctcgagacacacacgctcaactggtttatgcgctctgagctcaaaagttggac 6615
DB 4015 GTTGGTGTTCGAGACACACACGCTCACGTGTTTATGCCGCTCTGAGCTCAGAGTTGGAC 4074
QY 6616 gataaggtcggcacatatattagcgcgctcacctcaatcagatccgagcgtttgctt 6675
DB 4075 GATAAGTCCGACATATATTAGCAGCGCTCGCCTCAATCAGTATCCGACGCTTTCCTT 4134
QY 6676 cacacgttgcggccggcgtggttcttccgaggtatga 6714
DB 4135 CACACGTTCCGCCCGGCTGCTGCTTTTCGAGGATATAGA 4173

RESULT 2

US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner

Qy	4667	ataacttgaagaggagagctgcggtctttttctcaaaagctgggtgaggggtataacctata	4726
Db	4307	ACAACCTTTTAAGCGCAAGCGTCTCTGTGCCTGAAAGCCTGACGATGAGAATATATCTG	4248
Qy	4727	atggtggtagcactgtttcatcagggtggactcgctgcctcagagagatactcttaacggcaa	4786
Db	4247	ATCGTGTGATGCATCTTGGCATGTGCTGCACAATAGCCAAAGAAAGTAGCAGAAG	4188
Qy	4787	t 4787	
Db	4187	T 4187	

RESULT 6

US-08-036-555B-157

; Sequence 157, Application US/08036555B

; Patent No. 5530109

; GENERAL INFORMATION:

; APPLICANT: Goodearl, Andrew; Stroobant, Paul;

; APPLICANT: Minghetti, Luisa; Waterfield, Michael; Marchioni, Mark;

; APPLICANT: Chen, Mao Su; Hiles, Ian

; TITLE OF INVENTION: Glial Mitogenic Factors, Their

; TITLE OF INVENTION: Preparation and Use

QY	2007	cacccttttcttatgtacgcaactctgcccgggaacgctgttttaagctagataggggata	206
Db	183	TATGAGAAAGCGGAGG 199	122
QY	2067	ccgttggtgttaccatcatctcactatgtgtaggggtcagaagctccctggaactcgaggt	2126
Db	123	TGTGTATCGTCTGCCAAACTACGTAATGCGCAGCTTCTTACAAGCATCTTGGGATGAATT	182
QY	2127	tacgtataggtcgttgg 2143	
Db	183	TATGAGAAAGCGGAGG 199	
RESULT	9		
US-08-469-526A-157			
Sequence 157, Application US/08469526A			
Patent No. 5792849			
GENERAL INFORMATION:			
APPLICANT: Goodearl, Andrew			
APPLICANT: Stroobant, Paul			
APPLICANT: Minghetti, Luisa			
APPLICANT: Waterfield, Michael			
APPLICANT: Marchionni, Mark			
APPLICANT: Chen, Maio Su			
APPLICANT: Hiles, Ian			
TITLE OF INVENTION: GLIAL MITOGENIC FACTORS, THEIR			
MEDIUM TYPE: DISKETTE			
OPERATING SYSTEM: PC-DOS			
SOFTWARE: Wordperfect			
CURRENT APPLICATION DATA:			
APPLICATION NUMBER: US/08/249,322A			
FILING DATE: 26-MAY-1994			
CLASSIFICATION: 435			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER: 08/036,555			
FILING DATE: 24-MAR-1993			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER: 07/965,173			
FILING DATE: 23-OCT-1992			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER: 07/940,389			
FILING DATE: 03-SEP-1992			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER: 07/907,138			
FILING DATE: 30-JUN-1992			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER: 07/863,703			
FILING DATE: 03-APRIL-1992			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER: U.K. 91 07566.3			
FILING DATE: 10-APRIL-1991			
ATTORNEY/AGENT INFORMATION:			
NAME: Tsai, Christine H.			
REGISTRATION NUMBER: 34,266			
REFERENCE/DOCKET NUMBER: LUD 250.4			
TELEPHONE: (212) 688-9200			
TELEFAX: (212) 838-3884			
INFORMATION FOR SEQ ID NO: 157:			
SEQUENCE CHARACTERISTICS:			
LENGTH: 210			
TYPE: nucleic acid			
STRANDEDNESS: single			
TOPOLOGY: linear			
US-08-249-322A-157			
Query Match	0.58	Score 35.4	DB 1
Best Local Similarity	48.7%	Pred. No. 0.14	
Matches	96	Conservative	0
		Mismatches	101
		Indels	0
		Gaps	0
QY	1947	caatgttggtacgtgtgtgagatgtatgaacatctctctcactgaacggtgtttt	2006
Db	3	CCATCTTGTCTGACAGTGTGACAGAGGAGAAACTTCTGTGTGATGGAGGCGAGTGTCT	62

```

Query Match          0.5%; Score 35.4; DB 1; Length 210;
Best Local Similarity 48.7%; Pred. No. 0.14;
Matches 96; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 1947 caatgttggtcagtggtgagatgtatgaacattcctctcttaacgtaagcggttttt 2006
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3 CCATCTTGTCAAGTGTGCAGAGAGAGAAAAACTTTCTGTGTGAATGGAGGCGGTGCTT 62

QY 2007 cacccttttctattacgcactctgcgggaacgtgtttaagctagagtagagggata 2066
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 63 CATGGTGAAGACCTTTCAAATCCCTCAAGATACCTTGCAAGTCCCAAAATGAGTTTAC 122

QY 2067 ccggttggtgtaccatcatctcactatgtgttaggctcagaagtcacctgggaactgaggt 2126
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 123 TGGTGATCGCTGCCAAAACACTAGTAATGCCACGCTTCTACAAGCATCTTGGGATTGAATT 182

QY 2127 tacgtataggtcgttgg 2143
    ||||| ||||| |||||
Db 183 TATGGAGAAAGCGGAGG 199

RESULT 10
US-08-734-591A-157
; Sequence 157, Application US/08734591A
; Patent No. 5854220
; GENERAL INFORMATION:
; APPLICANT: Goodearl, Andrew
; APPLICANT: Stroobant, Paul
; APPLICANT: Minghetti, Luisa
; APPLICANT: Waterfield, Michael
; APPLICANT: Hiles, Ian
; APPLICANT: Marchionni, Mark
; APPLICANT: Chen, Mario
; TITLE OF INVENTION: GLIAL MITOGENIC FACTORS, THEIR
; TITLE OF INVENTION: PREPARATION AND USE
; NUMBER OF SEQUENCES: 187
; CORRESPONDENCE ADDRESS:
; STREET: 176 Federal Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM Compatible Pentium
; OPERATING SYSTEM: Windows95
; SOFTWARE: Wordperfect (Version 7.0)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/734,591A
; FILING DATE: 22-OCT-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/470,335
; FILING DATE: 06-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/036,555
; FILING DATE: 03-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/965,173
; FILING DATE: 23-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/940,389
; FILING DATE: 03-SEP-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/907,138
; FILING DATE: 30-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/863,703
; FILING DATE: 03-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: UK 91 07566.3
; FILING DATE: 10-APR-1991

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Bieker-Brady, Kristina
; REGISTRATION NUMBER: 39,109
; REFERENCE/DOCKET NUMBER: 04585/00200P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 428-0200
; TELEFAX: (617) 428-7045
; TELEX:
; INFORMATION FOR SEQ ID NO: 157:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 210
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-734-591A-157

Query Match          0.5%; Score 35.4; DB 2; Length 210;
Best Local Similarity 48.7%; Pred. No. 0.14;
Matches 96; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 1947 caatgttggtcagtggtgagatgtatgaacattcctctcttaacgtaagcggttttt 2006
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3 CCATCTTGTCAAGTGTGCAGAGAGAGAAAAACTTTCTGTGTGAATGGAGGCGGTGCTT 62

QY 2007 cacccttttctattacgcactctgcgggaacgtgtttaagctagagtagagggata 2066
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 63 CATGGTGAAGACCTTTCAAATCCCTCAAGATACCTTGCAAGTCCCAAAATGAGTTTAC 122

QY 2067 ccggttggtgtaccatcatctcactatgtgttaggctcagaagtcacctgggaactgaggt 2126
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 123 TGGTGATCGCTGCCAAAACACTAGTAATGCCACGCTTCTACAAGCATCTTGGGATTGAATT 182

QY 2127 tacgtataggtcgttgg 2143
    ||||| ||||| |||||
Db 183 TATGGAGAAAGCGGAGG 199

RESULT 11
US-08-469-660-157
; Sequence 157, Application US/08469660
; Patent No. 5876973
; GENERAL INFORMATION:
; APPLICANT: Gwynne, David I.; Marchionni, Mark;
; APPLICANT: McBurney, Robert N.
; TITLE OF INVENTION: INHIBITORS OF CELL PROLIFERATION,
; TITLE OF INVENTION: THEIR PREPARATION AND USE
; NUMBER OF SEQUENCES: 184
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; ZIP: 0211-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,660
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/011,396
; FILING DATE: 29-JAN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/984,085
; FILING DATE: 01-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/951,747
; FILING DATE: 25-SEP-1992
; PRIOR APPLICATION DATA:

```



```

; TITLE OF INVENTION: METHODS FOR ALTERING
; TITLE OF INVENTION: MUSCLE CONDITION
; NUMBER OF SEQUENCES: 185
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360
; MEDIUM TYPE: kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/05083C
; FILING DATE: 06-MAY-94
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/209,204
; FILING DATE: 08-MAR-94
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/059,022
; FILING DATE: 06-MAY-93
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 04585/028W01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 153:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 210
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; PCT-US94-05083C-153

```

	Query Match	0.5%	Score 35.4;	DB 4;	Length 210;
	Best Local Similarity	48.7%;	Pred. No.	0.14;	
	Matches 96;	Conservative	0;	Mismatches 101;	Indels 0; Gaps 0;
Qy	1947	caatgttggtcagvtggtgagtatgaacattcttctaagcgtaagcgagtttt	2006		
Db	3	CCATCTTGTCAAGTGTGCAGACAGGAGAAACTTTCTGTGTGAATGGAGCGGAGTGCTT	62		
Qy	2007	caccttttcttatgcagcacttcgtccggaaacggttttaagctagatgtagaggata	2066		
Db	63	CATGGTAAAGACCTTTCMAATCCCTCAAGATACTTGTGCAAGTGCCCAAATGAGTTTAC	122		
Qy	2067	ccgtgtggttaccatcatctcactatgtgtagggctcagaagtcacctgaaactgaagt	2126		
Db	123	TGGTGATCGCTGCCAANAACATACGTAATGCCAGCTTCTACAAGCATCTTGGGATTGAAT	182		
Qy	2127	tacgtatataggtcgttgg	2143		
Db	183	TATGGAGAAGCGCAGG	199		

RESULT 15
PCT-US95-06846A-157
; Sequence 157, Application PC/TUS9506846A
; GENERAL INFORMATION:
; APPLICANT: Goodearl, Andrew David; Stroobant, Paul;
; APPLICANT: Minghetti, Luisa; Waterfield, Michael; Marchioni, Mark;
; APPLICANT: Chen, Mao Su; Hiles, Ian
; TITLE OF INVENTION: Glial Mitogenic Factors, Their
; TITLE OF INVENTION: Preparation and Use
; NUMBER OF SEQUENCES: 178

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/06846A
FILING DATE: 25-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/249,322
FILING DATE: 26-MAY-1994
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/036,555
FILING DATE: 24-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/965,173
FILING DATE: 23-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/940,389
FILING DATE: 03-SEP-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/907,138
FILING DATE: 30-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/863,703
FILING DATE: 03-APRIL-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.K. 91 07566.3
FILING DATE: 10-APRIL-1991
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, Norman D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5250.5
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 157:
SEQUENCE CHARACTERISTICS:
LENGTH: 210
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
PCT-US95-06846A-157

```

	Query Match	0.5%	Score 35.4;	DB 4;	Length 210;
	Best Local Similarity	48.7%;	Pred. No. 0.14;		
	Matches 96;	Conservative 0;	Mismatches 101;	Indels 0;	Gaps 0;
Qy 1947	caatgttggtcagtggtgtagatgtatgaacattcttctctaactgaacgggtttt	2006			
Db 3	CCATCTTGTCAAGTGTGCAGAGAGGAAACATTCTGTGTAATGAGCGGAGTGCCTT	62			
Qy 2007	cacottttcttatgacgcacttcgtccgggaacggttttaagctagatgtagagggata	2066			
Db 63	CATGGTGAAGACCTTTCAAATCCCTCAAGATACTTGTGCAAGTGCCCAATGAGTTTAC	122			
Qy 2067	ccgttgtggtttaccatcatctcactatgtgttagggtccagaagtcacactgaacctgaggt	2126			
Db 123	TGGTGATCGCTGCCAAACATACGTAATGGCCAGCTTCTACAGCATCTTGGGATTTGAAT	182			
Qy 2127	taogtataggtcgttgg	2143			
Db 183	TATGGAGAAACGGGAG	199			

Search completed: February 14, 2001, 23:29:54
Job time: 23723 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 14, 2001, 16:18:27 ; Search time 3478.09 Seconds
(without alignments)
13527.043 Million cell updates/sec

Title: US-09-301-906-3
Perfect score: 6714
Sequence: 1 atggactacattgccccatt.....gttgcttcagggtatatga 6714

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_estba:*
42: em_estfun:*
43: em_esthum1:*

44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estov1:*
68: em_estov2:*
69: em_estpl1:*
70: em_estpl2:*
71: em_estpl3:*
72: em_estpl4:*
73: em_estpl5:*
74: em_estrol:*
75: em_estro2:*
76: em_estro3:*
77: em_estro4:*
78: em_estro5:*
79: em_estro6:*
80: em_estro7:*
81: em_estro8:*
82: em_estro9:*
83: em_estro10:*
84: em_estro11:*
85: em_estro12:*
86: em_estro13:*
87: gb_est41:*
88: gb_est42:*
89: gb_est43:*
90: gb_est44:*
91: gb_est45:*
92: gb_est46:*
93: gb_est47:*
94: gb_est48:*
95: gb_est49:*
96: gb_est50:*
97: gb_est51:*
98: gb_est52:*
99: gb_est53:*
100: gb_est54:*
101: gb_est55:*
102: gb_est56:*
103: gb_est57:*
104: gb_est67:*
105: gb_est68:*
106: gb_est69:*
107: gb_est70:*
108: gb_est71:*
109: gb_est72:*
110: gb_est73:*
111: gb_est74:*
112: em_esthum21:*
113: em_esthum22:*
114: em_esthum23:*
115: em_estom1:*
116: em_estom2:*

117: em_estpl6:*
118: em_estpl7:*
119: em_estpl8:*
120: em_estpl9:*
121: em_estpl10:*
122: em_estpl11:*
123: em_estpl12:*
124: em_estpl13:*
125: em_estpl14:*
126: em_estpl15:*
127: em_estpl16:*
128: em_estpl17:*
129: em_estpl18:*
130: em_estpl19:*
131: em_estpl20:*
132: em_estpl21:*
133: em_estpl22:*
134: em_estpl23:*
135: em_estpl24:*
136: em_estpl25:*
137: em_estpl26:*
138: em_estpl27:*
139: em_estpl28:*
140: em_estpl29:*
141: em_estpl30:*
142: em_estpl31:*
143: em_estpl32:*
144: em_estpl33:*
145: em_estpl34:*
146: em_estpl35:*
147: em_estpl36:*
148: em_estpl37:*
149: em_estpl38:*
150: em_estpl39:*
151: em_estpl40:*
152: em_estpl41:*
153: em_estpl42:*
154: em_estpl43:*
155: em_estpl44:*
156: em_estpl45:*
157: em_estpl46:*
158: em_estpl47:*
159: em_estpl48:*
160: em_estpl49:*
161: em_estpl50:*
162: em_estpl51:*
163: em_estpl52:*
164: em_estpl53:*
165: em_estpl54:*
166: em_estpl55:*
167: em_estpl56:*
168: em_estpl57:*
169: em_estpl58:*
170: em_estpl59:*
171: em_estpl60:*
172: em_estpl61:*
173: em_estpl62:*
174: em_estpl63:*
175: em_estpl64:*
176: em_estpl65:*
177: em_estpl66:*
178: em_estpl67:*
179: em_estpl68:*
180: em_estpl69:*
181: em_estpl70:*
182: em_estpl71:*
183: em_estpl72:*
184: em_estpl73:*
185: em_estpl74:*
186: em_estpl75:*
187: em_estpl76:*
188: em_estpl77:*
189: em_estpl78:*

190: gb_gss25:*
191: gb_gss26:*
192: gb_gss27:*
193: gb_gss28:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	50.6	0.8	346	21	AI542359	AI542359 SD08652.5
2	50.6	0.8	448	21	AI542833	AI542833 SD09311.5
3	50.6	0.8	573	21	AI542293	AI542293 SD08356.5
4	50.6	0.8	580	21	AI543590	AI543590 SD10387.5
5	50.6	0.8	580	21	AI546664	AI546664 SD09861.5
6	50.6	0.8	589	21	AI546576	AI546576 SD06393.5
7	50.6	0.8	629	21	AI531540	AI531540 SD02626.5
8	47.4	0.7	636	21	AI532043	AI532043 SD03401.5
9	47.4	0.7	592	21	AI542434	AI542434 SD08761.5
10	46	0.7	1101	190	CNS000D1	AL065414 Drosophila
11	45.8	0.7	1101	190	CNS0017Y0	AL108642 Drosophila
12	45.4	0.7	482	21	AI531508	AI531508 SD02578.5
13	44.2	0.7	997	190	CNS005TE	AL060767 Drosophila
14	42.2	0.6	1101	190	CNS0039G	AL063921 Drosophila
15	41.6	0.6	726	176	AZ196050	AZ196050 SP_1031_A
16	40.8	0.6	583	27	AI995779	AI995779 701548343
17	40.6	0.6	1101	190	CNS00L42	AL078714 Drosophila
18	40	0.6	556	27	AI995192	AI995192 701502553
19	40	0.6	990	190	CNS00LRW	AL078672 Drosophila
20	39.8	0.6	436	145	T43526	T43526 6789 Lambda
21	39.8	0.6	939	190	CNS00CNG	AL059400 Drosophila
22	39.4	0.6	920	190	CNS0062R	AL061710 Drosophila
23	39.4	0.6	1101	190	CNS016LW	AL106910 Drosophila
24	39.2	0.6	390	13	AA861672	AA861672 ak20h04.s
25	39.2	0.6	412	17	AI163552	AI163552 A044P300
26	39.2	0.6	511	93	AW659211	AW659211 96177 MAR
27	38.8	0.6	987	190	CNS00418	AL066537 Drosophila
28	38.6	0.6	943	190	CNS0021M	AL097720 Drosophila
29	38.6	0.6	1101	190	CNS00E1C	AL067660 Drosophila
30	38.4	0.6	401	23	AI696977	AI696977 wg76d11.x
31	38.4	0.6	451	20	AI418793	AI418793 tg33e01.x
32	38.4	0.6	488	17	AI160476	AI160476 qb9b06.x
33	38.4	0.6	523	24	AI755061	AI755061 cr34h08.x
34	38.2	0.6	925	190	CNS001BN	AL074475 Drosophila
35	38	0.6	631	139	C22941	C22941 C22941 Dict
36	38	0.6	673	29	AU060170	AU060845 AU060845
37	38	0.6	754	29	AU060170	AU060170 AU060170
38	38	0.6	1101	190	CNS00KK2	AL077673 Drosophila
39	38	0.6	1370	135	BE729475	BE729475 601564834
40	37.6	0.6	398	19	AI342750	AI342750 qo26c09.x
41	37.6	0.6	428	23	AI659489	AI659489 tx34g04.x
42	37.6	0.6	436	23	AI680779	AI680779 tx34g04.x
43	37.6	0.6	444	16	AI096769	AI096769 qa06h05.x
44	37.6	0.6	447	18	AI275580	AI275580 ql77c04.x
45	37.6	0.6	449	87	AW207125	AW207125 UI-H-B11-

ALIGNMENTS

RESULT 1

AI542359 LOCUS AI542359 346 bp mRNA EST 22-MAR-1999
DEFINITION SD08652.5prime SD Drosophila melanogaster Schneider L2 cell culture
pot2 Drosophila melanogaster cDNA clone SD08652 5prime, mRNA
sequence.
ACCESSION AI542359
VERSION AI542359.1 GI:4459732
KEYWORDS EST.

```

SOURCE          fruit fly.
ORGANISM        Drosophila melanogaster
                Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
                Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
                Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE       1 (bases 1 to 346)
AUTHORS         Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C., Brokstein
                ,P., Lewis,S. and Rubin,G.M.
TITLE           BDP/HHMI Drosophila EST Project
JOURNAL         Unpublished (1997)
COMMENT         Contact: Harvey, D.
                G. M. Rubin-Molecular and Cell Biology
                University of California Berkeley
                539 USA, Berkeley, CA 94720-3200, USA
                Fax: 510 643 9947
                Email: http://www.fruitfly.org/EST\_est@fruitfly.berkeley.edu
                Plate: 86 row: E column: 4
                High quality sequence stop: 292.
FEATURES        location/Qualifiers
                source
                1..346
                /organism="Drosophila melanogaster"
                /db_xref="taxon:7227"
                /clone="SD08652"
                /clone_lib="SD Drosophila melanogaster Schneider L2 cell
                culture pot2"
                /lab_host="DH5-alpha"
                /note="Vector: pot2; Site_1: EcoRI; Site_2: XhoI; Sized
                fractionated cDNAs were directly ligated into pot2.
                Plasmid cDNA library."
BASE COUNT      92 a 66 c 82 g 106 t
ORIGIN

Query Match      0.8%; Score 50.6; DB 21; Length 346;
Best Local Similarity 52.1%; Pred. No. 0.00059;
Matches 136; Conservative 0; Mismatches 124; Indels 1; Gaps

Qy 6011 aaagggtttggtgaggtgcacatgatgcataaggtcttactacaactaggcgctct 6070
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21 ATAGTATCTTTATTGACGAAGCCTTGATGTCCTCATCCAGGAGCTGTTTATGCTGTTT 80

Qy 6071 tcgcaacggcgcgctcggaaggcctctttttggagacataaatcagataccattcataa 6130
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 81 GGATGCTCGGTGCGTCAAAAAGTTAGATTTCCTGCTGACAACTTGCAAAATTCATTTGTA 140

Qy 6131 acagggaagaggtgtttagatgaggtgtcgttttttccaaagaaggaaagcggttg 6190
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 141 ACCGTCTCCGCAGCTTCAAGATGAGTTACCATCGTTGGAAAAAATGGTCCCAATGTGCG 200

Qy 6191 tatacactcttaaatcgacaggtgtccgttagatgttgcactgttgc-tcaatg 6249
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 201 GTTCACITTTTATCTCGTACCGCTGCTCTCGATGCTGTTATAGTTTACCGAGGACT 260

Qy 6250 accgtaaggggaacggaaaag 6270
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 261 ACCGTGAAGTGAACAAAAAAG 281

RESULT 2
AI542833
LOCUS          AI542833 448 bp mRNA EST 22-MAR-1999
DEFINITION    SD09311.5prime SD Drosophila melanogaster Schneider L2 cell culture
                pot2 Drosophila melanogaster cDNA clone SD09311 5prime, mRNA
                sequence.
ACCESSION     AI542833
VERSION       AI542833..1 GI:4460206
KEYWORDS      EST.
SOURCE        fruit fly.
ORGANISM      Drosophila melanogaster
                Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
                Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
                Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE     1 (bases 1 to 448)

```

AUTHORS Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C., Brokstein
P., Lewis,S. and Rubin,G.M.
TITLE BGP/HMI Drosophila EST Project
JOURNAL Unpublished (1997)
COMMENT Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
Plate: 93 row: A column: 11
High quality sequence stop: 340.
Location/Qualifiers
1..448
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="SD09311"
/clone_lib="SD Drosophila melanogaster Schneider L2 cell
culture pot2"
/lab_host="DH5-alpha"
/note="vector: pot2; Site.1: EcoRI; Site.2: XhoI: Sized
fractionated cDNAs were directly ligated into pOT2.
Plasmid cDNA library."
BASE COUNT 131 a 95 c 94 g 128 t
ORIGIN

Query Match 0.88; Score 50.6; DB 21; Length 448;
Best Local Similarity 52.18; Pred. No. 0.00066;
Matches 136; Conservative 0; Mismatches 124; Indels 1; Gaps 1;

Qy 6011 aaagggttttggtagtgagtgccatgatgcataaggtcttactacaactagcgctct 6070
Db 165 ATAGTATCTTTATTCACGAAGCCCTTGATGTCCATCCAGGAGCTGTTTTATGCTGTT 224
Qy 6071 tcgcacccggcgctgcggaagccctctttttggagacataaatcagataccattcataa 6130
Db 225 GGATGTCGTGTCGGTCAAAAGTTAGATTCTTCGGTGACAACTTGCAAATTCATTGTGTA 284
Qy 6131 acaggagaaagtgtttagtaggatgctgtctgttttttccaaagaagaagcgttg 6190
Db 285 ACCGTTCTCCGCAGTTCGAAGATGAGTTCAGCATCGTTTGGAAAAAATGGTCCAAATTCGC 344
Qy 6191 tatacaactctaaatcgtcacaggtgtccgttagatgtttgctactgttgtcc-tcaatg 6249
Db 345 GTTCACTTTTATCTGTCACGCTGCTCTGTCGATGCGTTTATAGTTTACCGACGACT 404
Qy 6250 accgtaaggggaaacgaaaaag 6270
Db 405 ACCGTGAAGTGAACAAAAAAG 425

RESULT 3
AI542293
LOCUS AI542293 573 bp mRNA EST 22-MAR-1999
DEFINITION SD08556.5prime SD Drosophila melanogaster Schneider L2 cell culture
pot2 Drosophila melanogaster cDNA clone SD08556 5prime, mRNA
sequence.
ACCESSION AI542293
VERSION AI542293.1 GI:4459666
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 573)
Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C., Brokstein
P., Lewis,S. and Rubin,G.M.
REFERENCE BGP/HMI Drosophila EST Project
AUTHORS Unpublished (1997)
TITLE Contact: Harvey, D.
JOURNAL G. M. Rubin-Molecular and Cell Biology
COMMENT


```
culture pot2"
/lab_host="DH5-alpha"
/notes=Vector: pot2; Site_1: ECORI; Site_2: XhoI; Sized
fractionated cDNAs were directly ligated into pot2.
Plasmid cDNA library."
BASE COUNT      167 a   124 c   122 g   167 t
ORIGIN

Query Match      0.8%; Score 50.6; DB 21; Length 580;
Best Local Similarity 52.1%; Pred. No. 0.00074;
Matches 136; Conservative 0; Mismatches 124; Indels 1; Gaps 1;

QY 6011 aaagggttttggatgaggtgcacatgatcatcaaggtctactacaactaggcgctct 6070
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12 ATAGTATCTTTATTGACGAAGCCTTGATGTCCTCCATCCAGGAGCTGTTTTATGCTGTTT 71
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 6071 tcgcaaccgagcgctcggaaggcctctttttgagacataataatcagataccattcataa 6130
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 72 GGATGCTGTGTCGCTCAAAAGTTAGATTCTCGTGACAACTTGCAAAATTCCTCAATTGTAA 131
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 6131 acaggggagaaaggtttagatgattgtctgtttttgttccaaagaagaagcggttg 6190
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 132 ACCGTTCTCCGCAGTTCAAGATGAGTACGATCGTTTGGAAAAAATGGTCCCAATTGTCG 191
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 6191 tatacacttctaatacgacagtgctcgttagatgttctgtactgtgtgtcc-tcaatg 6249
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 192 GTTCACCTTTTATCTGTCGTCGCTGCTCGATGTCGTTTATAGTTTACCGAGCACT 251
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 6250 accgtaagggaagcaaaaaaag 6270
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 252 ACCGTGAAGTGAACAAAAAAG 272
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 6
AI546576      589 bp      mRNA      EST      22-MAR-1999
LOCUS      SD06393.5prime SD Drosophila melanogaster Schneider L2 cell culture
DEFINITION      pot2 Drosophila melanogaster cDNA clone SD06393 5prime, mRNA
sequence.
ACCESSION      AI546576
VERSION      AI546576.1 GI:4463949
KEYWORDS      EST.
SOURCE      fruit fly.
ORGANISM      Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE      1 (bases 1 to 589)
AUTHORS      Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C., Brokstein
,P., Lewis,S. and Rubin,G.M.
TITLE      BDGP/HHMI Drosophila EST Project
JOURNAL      Unpublished (1997)
COMMENT      Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
High quality sequence stop: 456.
Location/Qualifiers
1..589
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="SD06393"
/clone_lib="SD Drosophila melanogaster Schneider L2 cell
culture pot2"
/lab_host="DH5-alpha"
/notes=Vector: pot2; Site_1: ECORI; Site_2: XhoI; Sized
fractionated cDNAs were directly ligated into pot2.
Plasmid cDNA library."
BASE COUNT      174 a   125 c   126 g   164 t

culture pot2"
/lab_host="DH5-alpha"
/notes=Vector: pot2; Site_1: ECORI; Site_2: XhoI; Sized
fractionated cDNAs were directly ligated into pot2.
Plasmid cDNA library."
BASE COUNT      167 a   124 c   122 g   167 t
ORIGIN

Query Match      0.8%; Score 50.6; DB 21; Length 589;
Best Local Similarity 52.1%; Pred. No. 0.00075;
Matches 136; Conservative 0; Mismatches 124; Indels 1; Gaps 1;

QY 6011 aaagggttttggatgaggtgcacatgatcatcaaggtctactacaactaggcgctct 6070
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 163 ATAGTATCTTTATTGACGAAGCCTTGATGTCCTCCATCCAGGAGCTGTTTTATGCTGTTT 222
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 6071 tcgcaaccgagcgctcggaaggcctctttttgagacataataatcagataccattcataa 6130
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 223 GGATGCTGTGTCGCTCAAAAGTTAGATTCTCGTGACAACTTGCAAAATTCCTCAATTGTAA 282
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 6131 acaggggagaaaggtttagatgattgtctgtttttgttccaaagaagaagcggttg 6190
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 283 ACCGTTCTCCGCAGTTCAAGATGAGTACGATCGTTTGGAAAAAATGGTCCCAATTGTCG 342
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 6191 tatacacttctaatacgacagtgctcgttagatgttctgtactgtgtgtcc-tcaatg 6249
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 343 GTTCACCTTTTATCTGTCGTCGCTGCTCGATGTCGTTTATAGTTTACCGAGCACT 402
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 6250 accgtaagggaagcaaaaaaag 6270
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 403 ACCGTGAAGTGAACAAAAAAG 423
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
AI531540      629 bp      mRNA      EST      18-MAR-1999
LOCUS      SD02626.5prime SD Drosophila melanogaster Schneider L2 cell culture
DEFINITION      pot2 Drosophila melanogaster cDNA clone SD02626 5prime, mRNA
sequence.
ACCESSION      AI531540
VERSION      AI531540.1 GI:4445675
KEYWORDS      EST.
SOURCE      fruit fly.
ORGANISM      Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE      1 (bases 1 to 629)
AUTHORS      Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C., Brokstein
,P., Lewis,S. and Rubin,G.M.
TITLE      BDGP/HHMI Drosophila EST Project
JOURNAL      Unpublished (1997)
COMMENT      Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
High quality sequence stop: 463.
Location/Qualifiers
1..629
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="SD02626"
/clone_lib="SD Drosophila melanogaster Schneider L2 cell
culture pot2"
/lab_host="DH5-alpha"
/notes=Vector: pot2; Site_1: ECORI; Site_2: XhoI; Sized
fractionated cDNAs were directly ligated into pot2.
Plasmid cDNA library."
BASE COUNT      188 a   132 c   133 g   176 t

culture pot2"
/lab_host="DH5-alpha"
/notes=Vector: pot2; Site_1: ECORI; Site_2: XhoI; Sized
fractionated cDNAs were directly ligated into pot2.
Plasmid cDNA library."
BASE COUNT      167 a   124 c   122 g   167 t
ORIGIN

Query Match      0.8%; Score 50.6; DB 21; Length 629;
Best Local Similarity 52.1%; Pred. No. 0.00077;
Matches 136; Conservative 0; Mismatches 124; Indels 1; Gaps 1;
```

```

QY 6011 aaagggttttggatgaggtgcacatgatgcatacgaagcttactacaactaggcgctct 6070
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 191 ATAGTATCTTTATTGACGAAGCCTTGATGTCCCATCCAGAGCTGTTTTTTATGCTGTTT 250
QY 6071 tcgcacaacggcgctgcgaagccctcttttttggagacataaatacagataccattcataa 6130
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 251 GGATGCTCTGGTCCGTCACAAAGTTAGATTTCTCGGTGACAACTTGCAAAATTCATTTGTA 310
QY 6131 acagggagaagtggttttaggagtgattgctgttttcttcttccaaagaagaagcgttg 6190
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 311 ACCGTTCTCCGAGTTCAAGATGAGTACGATCGTTTGAAAAAATGGTCCCAATGTGCG 370
QY 6191 tatacacttctaaacgtacaggtgctcggttagatgtttgctacttctgttgc-tcaatg 6249
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 371 GTTCACCTTTTATCTCGTACCGCTGCTCGTGCATGTCGTTTATAGGTTTACCGAGGACT 430
QY 6250 accgtaaggggaacggaaaaag 6270
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 431 ACCGTGAAGTGAACAAAAAG 451

RESULT 8
LOCUS AI532043 636 bp mRNA EST 18-MAR-1999
DEFINITION SD03401.5prime SD Drosophila melanogaster Schneider L2 cell culture
           pOT2 Drosophila melanogaster cDNA clone SD03401 5prime, mRNA
           sequence.
ACCESSION AI532043
VERSION AI532043.1 GI:4446178
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
           Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
           Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
           Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 636)
AUTHORS Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C., Brokstein
          ,P., Lewis,S. and Rubin,G.M.
TITLE BDGP/HMI Drosophila EST Project
JOURNAL Unpublished (1997)
COMMENT Contact: Harvey, D.
          G. M. Rubin-Molecular and Cell Biology
          University of California Berkeley
          539 LSA, Berkeley, CA 94720-3200, USA
          Fax: 510 643 9947
          Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
          Plate: 34 row: A column: 1
          High quality sequence stop: 505.
          Location/Qualifiers
FEATURES
    source
        1..636
            /organism="Drosophila melanogaster"
            /db_xref="taxon:7227"
            /clone_lib="SD03401"
            /clone_lib="SD Drosophila melanogaster Schneider L2 cell
            culture pOT2"
            /lab_host="DH5-alpha"
            /note="vector: pOT2; Site_1: EcoRI; Site_2: XhoI; Sized
            fractionated cDNAs were directly ligated into pOT2.
            Plasmid cDNA library."
BASE COUNT 189 a 145 c 127 g 175 t
ORIGIN

Query Match 0.7%; Score 47.4; DB 21; Length 636;
Best Local Similarity 51.2%; Pred. No. 0.0077;
Matches 111; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 6011 aaagggttttggatgaggtgcacatgatgcatacgaagcttactacaactaggcgctct 6070
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 414 ATAGTATCTTTATTGACGAAGCCTTGATGTCCTCCATCCAGAGCTGTTTTTTATGCTGTTT 473
QY 6071 tcgcacaacggcgctgcgaagccctcttttttggagacataaatacagataccattcataa 6130
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 474 GGATGCTCTGGTCCGTCACAAAGTTAGATTTCTCGGTGACAACTTGCAAAATTCATTTGTA 533
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 6131 acagggagaaggtgttagatgaggtgctgtgttttttccaaagaagaagcgttg 6190
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 534 ACCGTTCTCCGAGTTCAAGATGAGTACGATCGTTTGAAAAAATGGTCCCAATGTGCG 593
QY 6191 tatacacttctaaacgtacaggtgctcggttagatgt 6227
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 594 GTTCACCTTTTATCTCGTACCGCTGCTCGTGCATGT 630

RESULT 9
LOCUS AI542434 592 bp mRNA EST 22-MAR-1999
DEFINITION SD08761.5prime SD Drosophila melanogaster Schneider L2 cell culture
           pOT2 Drosophila melanogaster cDNA clone SD08761 5prime, mRNA
           sequence.
ACCESSION AI542434
VERSION AI542434
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
           Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
           Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
           Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 592)
AUTHORS Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C., Brokstein
          ,P., Lewis,S. and Rubin,G.M.
TITLE BDGP/HMI Drosophila EST Project
JOURNAL Unpublished (1997)
COMMENT Contact: Harvey, D.
          G. M. Rubin-Molecular and Cell Biology
          University of California Berkeley
          539 LSA, Berkeley, CA 94720-3200, USA
          Fax: 510 643 9947
          Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
          Plate: 87 row: F column: 1
          High quality sequence stop: 538.
          Location/Qualifiers
FEATURES
    source
        1..592
            /organism="Drosophila melanogaster"
            /db_xref="taxon:7227"
            /clone_lib="SD08761"
            /clone_lib="SD Drosophila melanogaster Schneider L2 cell
            culture pOT2"
            /lab_host="DH5-alpha"
            /note="vector: pOT2; Site_1: EcoRI; Site_2: XhoI; Sized
            fractionated cDNAs were directly ligated into pOT2.
            Plasmid cDNA library."
BASE COUNT 173 a 124 c 127 g 168 t
ORIGIN

Query Match 0.7%; Score 47; DB 21; Length 592;
Best Local Similarity 52.8%; Pred. No. 0.01;
Matches 124; Conservative 0; Mismatches 110; Indels 1; Gaps 1;

QY 6037 atgatgcataaggttactacaactaggctgttcgcgaacggcgctgcgaagcctc 6096
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12 ATGTCCCATCCAGAGCTGTTTTTTATGCTGTGTGGATGCTGTGGTCCGTCACAAAGTTAGA 71
QY 6097 tttttggagacataaatacagataccattcataaaggagagaaggtgttttaggaggtat 6156
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 72 TTTCTCGGTGACAACTTGCAAAATTCATTTGTAACCGTTCTCCGAGTTCAGATGAGT 131
QY 6157 tgtgctgttttggttccaaagaagaagcgttgtgtatacaacttctaaatcgtagaggtgt 6216
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 132 TACGATCGTTTGGAAAAAATGTCCTCAATGTCGGTTCATCTTTTATCTCGTACCGCTGT 191
QY 6217 caggttagatgtttgtacttctgtgttc-tcaatgaccgttaaggggaacggaaaaag 6270
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 192 CCGTGCATGTCGTTTATAGGTTTACCGACGACTACCGTGAAGTGAACAAAAAG 246

```

```
RESULT 10
CNS00001
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR01J16 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
AL065414
VERSION
AL065414.1 GI:4938827
KEYWORDS
GSS.
SOURCE
fruit fly.
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 1101)
AUTHORS
Direct Submission
JOURNAL
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mamoser in Piter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
source
1..1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR01J16"
/note="end : TET3"
BASE COUNT
280 a 104 c 123 g 211 t 383 others
ORIGIN
Query Match 0.7%; Score 46; DB 190; Length 1101;
Best Local Similarity 16.1%; Pred. No. 0.027;
Matches 72; Conservative 194; Mismatches 179; Indels 1; Gaps 1;
QY 1577 tgaactgtcatgtatgcaatccagctctacgttaagatgtgaagcgagagaatcaatg 1636
Db KRAATDDTATTTWKKTKAKKAGKAAKWKADRTKAWKAGKAGARGKATAKARAKRWK 711
QY 1637 agatcctcttcttccacagcgtggggagattcgtacgtgtccagtcactcttaactg 1696
Db DDKDTKKKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTK 771
QY 1697 aagcggttcaaatgtctgtcttactgttagtcagaaatcgcgaactgcgattctagag 1756
Db RTGARGDKDTGKKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTK 831
QY 1757 ccgatcggtttttatggttgatgtgtacgatataatccccgcgcagcagtagcagagcta 1816
Db DKADKKKAGKRWKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTK 890
QY 1817 tggataaagaggtgcgcgtggttttcgacatgctcttatgttcccgctggagttgtgt 1876
Db KKWKDDDDTAKDDWTKDTDWATAGTAGMDDDKAKDKDKDKDKDKDKDKDKDKDKDKDK 950
QY 1877 acggtaacggtgaagtcttacttgaagaactcgtacgttggtagaggggaaggtgatt 1936
Db 951 KGRKTAGAGDGKDKKKGKDKDKDRTDDAKATDGDGKDTWDDTTDTAKRRRRRRRGRGRDR 1010
QY 1937 acctggcctacaatgttggttcagtggtgagatgtgagacattctctcaacgtaa 1996
Db 1011 KGDKKKRWRARWRKGGKKTKDKDDGDKRWGRAKAKADAKADDDADAADAARAKAKA 1070
QY 1997 gcgggttttcacactcttcttatgta 2022
Db 1071 KDKDKAKKKADADDKDKDKAKARKDD 1096
RESULT 11
CNS017Y0/C
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence T7 end of BAC
BACN37H18 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
AL108642
VERSION
AL108642.1 GI:5628946
KEYWORDS
GSS.
SOURCE
fruit fly.
ORGANISM
Plasmid Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 1101)
AUTHORS
Direct Submission
JOURNAL
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11.
FEATURES
Location/Qualifiers
Source
1..1101
/organism="Drosophila melanogaster"
/plasmid="pBelobAC11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN37H18"
/note="end : T7"
BASE COUNT
167 a 155 c 2 g 671 t 106 others
ORIGIN
Query Match 0.7%; Score 45.8; DB 190; Length 1101;
Best Local Similarity 45.1%; Pred. No. 0.031;
Matches 116; Conservative 12; Mismatches 129; Indels 0; Gaps 0;
QY 6261 aacggaaaagtgttacccctgaaaaggtcgttagcggttaaggacaaaccagtagtaagtc 6320
Db AAAAAAAGAAATTRGGTAAAGATGGAGGAATGGGTGGGAGATAAAGTGGTATCAAT 720
QY 6321 gctgtccaaaggcccaattggaaccactgatgcgtagctgaaataaacctcagcgtgta 6380
Db 719 GGAATAAAGAAAGAGGATATAGAGAAATGAAGATGAAGAGAGGATTATWRTWAAGAGGTGAR 660
QY 6381 ctgtgcataccagctgtgagagctcgatatgaagagctcgttgagagaaagaaa 6440
Db 659 AAAGTTTGGAAATTAAGATRRATTAAGTAAGAAAGAAAGAAAGGAATTAAGGAATAGAMA 600
QY 6441 agaaacaccagtgatgacatgcatgaagcacacaggggaaaaacattcagtgatgtgggtatt 6500
Db 599 RRARAAGAAATRGATAGTAGTGAATGRAAAAAAAGGAGAAATATGATGTTGATTGAGGATT 540
QY 6501 gtttagacgaagaag 6517
```

```

Db 539 GGAATCAAAAGGAAG 523
| | | | | | | |
RESULT 12
A1531508
LOCUS
DEFINITION
SD02578.5prime SD Drosophila melanogaster Schneider L2 cell culture
pOT2 Drosophila melanogaster cDNA clone SD02578 5prime, mRNA
sequence.
A1531508
VERSION
A1531508.1 GI:4445643
EST.
SOURCE
fruit fly.
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 482)
AUTHORS
Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C., Brokstein
P., Lewis,S. and Rubin,G.M.
TITLE
BDGP/HIMI Drosophila EST Project
JOURNAL
Unpublished (1997)
COMMENT
Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: 25 row: G column: 6
High quality sequence stop: 447.
Location/Qualifiers
1. 482
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="SD02578"
/culture="pot2"
/lab_host="DH5-alpha"
/note="Vector: pOT2; Site 1: EcoRI; Site 2: XhoI; Sized
fractionated cDNAs were directly ligated into pOT2.
Plasmid cDNA library."
BASE COUNT
138 a 100 c 106 g 138 t
ORIGIN
Query Match 0.7%; Score 45.4; DB 21; Length 482;
Best Local Similarity 55.4%; Pred. No. 0.029;
Matches 108; Conservative 0; Mismatches 86; Indels 1; Gaps 1;
QY 6077 ccggcgcgcggaaggccctcttttggagacataaatacagataccattcataaacagg 6136
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 27 CTGGTCGGTCMAAAGTTAGATTCTCGGTGACAACTTGCAATTCATTTGTAAACCGTT 86
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 6137 agaagtggttaggattgctgtgttttggttccaaagaaggaaagcgtgtgtatataca 6196
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 87 CTCGCAGTTCAAGATGAGTTACGATCGTTGGAAAAATTGTCCTCAATTCGCGTTTAC 146
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 6197 cttctaaatcgatcaggttcggttagattgttgcacttgtgttc-tcaatgaccgta 6255
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 147 TTTTATCTCGTACCGCTGCTGTGATGTCGTTTATACCGACGACTACCGTG 206
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 6256 aggggaacggaag 6270
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 207 AAGTGACAAANAAG 221
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 13
CNS005TE/c
LOCUS
DEFINITION
CNS005TE 997 bp DNA GSS 03-JUN-1999
Drosophila melanogaster genome survey sequence TE3 end of BAC #
BAC12K22 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
AL063921
ACCESSION
AL063921.1 GI:4941778

```

```

AL060767
VERSION
AL060767.1 GI:4943573
GSS.
SOURCE
fruit fly.
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 997)
AUTHORS
Genoscope.
TITLE
Direct Submission
JOURNAL
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammos in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
Location/Qualifiers
1. 997
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/culture="BAC12K22"
/note="end : TET3"
BASE COUNT
89 a 99 c 13 g 258 t 538 others
ORIGIN
Query Match 0.7%; Score 44.2; DB 190; Length 997;
Best Local Similarity 21.8%; Pred. No. 0.095;
Matches 53; Conservative 88; Mismatches 102; Indels 0; Gaps 0;
QY 6282 aaagtcgttagcgttaagacaaccagtagtaaatcgctgtccaaagccaattgg 6341
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 978 ARRRRRGARGARARRRRRRARGARAGARAAARAGARAGARARRRRRRAG 919
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 6342 aaccactgatgactgacgtgaaataaacgctgactgtctgtgcatgccactgtga 6401
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 918 RRAGRAGRSAGGGRRGRRRAAARAARARARARARARRRRARARARGARRRR 859
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 6402 gaagtcggatatgaagaggtcgttgaaagggaaaaagaaagaaaccaccagtgtacagt 6461
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 858 GRRRRRGRAGRRRRARGAGRRGRGRRRRARRRARARARARRRRARARARR 799
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 6462 qcatgaagcacaggaaaaaacattcagatgtgtgtattgttagacgaagaaaccca 6521
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 798 RRAGRRRRARGARRRRRRGRRRRRRRRRRARRRRARARARARAFARRRRRR 739
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 6522 tga 6524
| |
Db 738 AGA 736
| |

RESULT 14
CNS0039G/c
LOCUS
DEFINITION
CNS0039G 1101 bp DNA GSS 03-JUN-1999
Drosophila melanogaster genome survey sequence TE3 end of BAC #
BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
AL063921
ACCESSION
AL063921.1 GI:4941778

```


KEYWORDS	GSS.				
SOURCE	fruit fly.				
ORGANISM	Drosophila melanogaster				
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;				
	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;				
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
REFERENCE	1 (bases 1 to 1101)				
AUTHORS	Genoscope.				
TITLE	Direct Submission				
JOURNAL	Submitted (02-JUN-1999) Genoscope - Centre National de Sequences ;				
	BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr				
	- Web : www.genoscope.cns.fr)				
COMMENT	Determination of this BAC-end sequence was carried out as part of a				
	collaboration with the Berkeley Drosophila Genome Project (BDGP).				
	The BDGP is constructing a physical map of the Drosophila				
	melanogaster genome using these BACs. For further information				
	please see http://www.fruitfly.org The BDGP Drosophila				
	melanogaster BAC library was prepared by Kazutoyo Osoegawa and				
	Aaron Mammoser in Pieter de Jong's laboratory in the Department of				
	Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,				
	NY. The library is named RPCI-98 and was constructed by partial				
	EcoRI digestion of Drosophila DNA provided by the BDGP from the				
	isogenic strain Y2; cn bw sp, the same strain used for the BDGP's				
	p1 and EST libraries. A more detailed description of the library				
	and how to order individual BAC clones, the entire library, or				
	filters for hybridization from the BACPAC Resource Center can be				
	found at http://bacpac.med.buffalo.edu/drosophila_bac.htm .				
FEATURES	Location/Qualifiers				
source	1..1101				
	/organism="Drosophila melanogaster"				
	/db_xref="taxon:7227"				
	/clone_lib="RPCI-98"				
	/clone="BACR08K10"				
	/note="end : TET3"				
BASE COUNT	201 a 64 c 131 g	202 t	503 others		
ORIGIN					

	Query Match	0.6%	Score 42.2;	DB 190;	Length 1101;
	Best Local Similarity	14.1%	Pred. No. 0.42;		
	Matches	62;	Conservative 210;	Mismatches 168;	Indels 1; Gaps 1;
Qy	5762	tgggtggcggaatcgtttaagcttccttgcgaataataagtctcataatgctccccccaggtagcg	5821		
Db	968	KDRKGGDDKKGGKKKAAKAAKWATKWMDWDWKDKWGAADKRDRAADDGAGDKDDD	909		
Qy	5822	gtaagacqacqacgttagtgagcaaatcgtttaagtcaccacaatgacacagccaccatta	5881		
Db	908	GKGKDADDTDGTKDDDDKDWDDWDAKAKGTGWGDATWAWAATDWWWGWAADNWMTWDA	849		
Qy	5882	cggctaattgtgggaagt-tctgagagcataaataatgcggtgaaagaagagatcccgaat	5940		
Db	848	ADWWADDWRDAWAKWDDAWAGARTADRRDWDGRAGKRGARKRKRDRKRAADKROAAD	789		
Qy	5941	ttsgaaggtctcaacagtgtcacacagtttaactccagggttgtaaactttatcgtcagg	6000		
Db	788	DRDDAATTTTWTTRDTDDWKNKTDTTWRAADRTWDRDDDDDRDRAGTAGRKRRTW	729		
Qy	6001	ggaatgtataaaaagggttttgttgtagtgaggtgcacatgatgcatacaggcttactcaa	6060		
Db	728	KRRWKRRTDRDDADADDTARDDRRRRGGDGADAGKGTGKRKRNRDRATWDRDTAWA	669		
Qy	6061	ctaggcgctctcgcaacgcgcgcgtcgaaagccctcttttggagacataaacagata	6120		
Db	668	DAAWTTTTDTDDWDDKDRRRKGAARRRRFTARAADWNWTKAWDWMKDWKTRADRWD	609		
Qy	6121	ccattcatcaaacaggagaaggttttaggatgatgtcgtgtttttgttcccaagaag	6180		
Db	608	RWAADTWDARKADRWAKARAWRARDRARAARDREWTTKGTTTATTWTAARAANW	549		
Qy	6181	gaagcgcttgatacattct	6201		
Db	548	AWAAWATTTATWTTTWTWT	528		

RESULT	15
AZ196050/c	
LOCUS	
DEFINITION	AZ196050 726 bp DNA GSS 31-AUG-2000 SP_1031.A2.E05.SP6E Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library Strongylocentrotus purpuratus genomic clone Plate-1031 Col-10 Row-I, DNA sequence.
ACCESSION	AZ196050
VERSION	AZ196050.1 GI:8389873
KEYWORDS	GSS.
SOURCE	Strongylocentrotus purpuratus.
ORGANISM	Strongylocentrotus purpuratus Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinoidea; Euchinoidea; Echinacea; Echinoidae; Strongylocentrotidae; Strongylocentrotus. 1 (bases 1 to 726)
REFERENCE	Smarton,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R., Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray ,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J, Davidson,E.H. and Hood,L.
AUTHORS	
TITLE	A sea urchin genome project: Sequence scan, virtual map, and additional resources
JOURNAL	Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)
MEDLINE	20402566
COMMENT	Contact: Cameron, RA, Davidson, EH, Hood, L Division of Biology 156-29 California Institute of Technology Pasadena California 91125, USA Tel: (626) 395-8421 Fax: (626) 793-3047 Email: acameron@caltech.edu Plate: 1031 row: I column: 10 Seq primer: SP6 Class: BAC ends High quality sequence stop: 726. Location/Qualifiers 1..726 /organism="Strongylocentrotus purpuratus" /db_xref="taxon:7668" /clones="Plate=1031 Col=10 Row=I" /clone_lib="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library" /note="Organ: sperm; Vector: BACE3.6; BAC Clones in B-Coli DH10B"
FEATURES	
source	253 a 210 c 64 g 199 t
BASE COUNT	
ORIGIN	
Query Match	0.68; Score 41.6; DB 176; Length 726;
Best Local Similarity	44.6%; Pred. No. 0.53;
Matches 164; Conservative	0; Mismatches 204; Indels 0; Gaps 0;
Qy 3766	gagggggaggaaagtgtgcacgatgagcgttagctaccgggaacttgactatgacgcaa 3825
Db 582	GATCAGATGATGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 523
Qy 3826	gggtcggtccatctgggaactcacgatgaagcgggttcccggtagcagaacgatgcc 3885
Db 522	GGTGTGATGATGATGGTGATGAAGGTGATGGTCATGATGGTCATGATTATGATGAGGAT 463
Qy 3886	acttcagtgtctcaagcatgatgttgtcacaaatgtgcgcgacgtggggattagcacc 3945
Db 462	GATCAGGATGATGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 403
Qy 3946	aacgggggaagttaactggtgaagaagagaccattccactcgaaagcgtgcaatacacttat 4005
Db 402	GCTGATGATTATGATGATTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 343
Qy 4006	gtcaggaagaagggttgcacctgtgcagctgtggcgaaaagacaaggtgatccccgcgggt 4065
Db 342	GGTGGTGATGATGATGATGCGHATGATGGTGTCTCATGCTGATGATGATGATGATGATGAT 283

	Query Match	0.6%	Score 41.6;	DB 176;	Length 726;
	Best Local Similarity	44.8%;	pred. No.	0.53;	
	Matches 164;	Conservative	0;	Mismatches 204;	Indels 0; Gaps 0;
Qy	3766	gagggagggaatgtgtacgatgagcacgttagttaccogggactttagctatacgcaa	3825		
Dd	582	GATCAGAATGATGGTGTATGATGATGATCATGTCGTGATGATGATGATGATGATGAT	523		
Qy	3926	ggtcctggtccattctgggacctcaacatgaagcgggtcccggtgacagataacogatggatcc	3885		
Dd	522	GGTGCTGATGATGATGGTTGATGAAGGTCATGGTCATGATGGTCATGATTATGATGAGCAT	463		
Qy	3886	acttcagtgtctcaagctatgatgttgtcacaaatgtgcgcacgtggggatttagaac	3945		
Dd	462	GATGAGGATGATGAGGATGATGATGATGATGATGTCGTGATGAAGGTGATGGTGTGATGAT	403		
Qy	3946	aacgggggaagttaactggttgagaagaagacaccattcaactcgaaagcgtgcgaatacacattat	4005		
Dd	402	GGTCATGATTATGATGATTATGATGATGATCATGACGATGCTGTGATGATGATGATTGCAT	343		
Qy	4006	gtcaggaagaaggttgccccogtctgcagctgtggcggaaaagacaaggtgatccctcggggt	4065		
Dd	342	GGTGCTGATGATGATGTCGTGATGGTGGTCTTCATGGTGTGATGATGATGATGATGATGAT	283		

```
QY 4066 tctgggtaccgctgacgctatggtctttttgtgaaagtgtgaaaaaagggtgctgacgatgtc 4125
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 282 GATAATGGTGTGATCACGGTGTGATGACGATGATGGTGTGATGATGATGACGATGATAAC 223
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 4126 tttcacca 4133
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 222 TTTCATCA 215
```

Search completed: February 14, 2001, 20:38:35
Job time: 13608 sec


```

ERYLDLPSPSPSDDDDHDMNDSANTNADGYRAVSFSYSGSSDVNDQKNADMESG
FIPPPVPDYDLRQNLPPTEKHLQIITRTSSFYSKHGQSEIVLRVKQGNPTFGELMP
DHLHLIYFRLVDHLELTKSVSEKKNESKDGALUSLGSVYGTVEDEEDANEESA
NDKSYSESAGDGVKVDNSNGPEGSKAATASKHSLPLNDHSAFIRKRNPSVAVNV
VFKQIOTIEKLVTSOKSOPKLEQIIVETPTTEMRKVIDKIVDFIQKNGELEATLVAQD
VKYGMFFPRLPSLHYAYRVKVLQAEELKSGDKGVIIRKEDVKOEKMGNAVKDASKHG
FGSVLPDDSAKREKLKMYSDPKPVELHNEPKPVOPOMRVNVYDANTAALQOASRGI
RNPGLGILTKPMDETSQTLGNDVSPSKSPDLAKSTGQSLGSGSTAASEADSSAEL
SEQLKARLAKAMFVAKLAPDAOPVOQAEPSSRSISVEPLDGLSIGLGAKAAREK
SSSIPIVASEKLADNGERRSKNRYSRSDRDEGKMEQGEESMDVEVTEFTKTD
KKHSCRRKHKTKYSKDRHSRDKHKHSESSDDEYHSRSRHRHRSKSDRHLKED
SSDGEHRHRSKSHKSDVDYKDKRSHHRSRKHKEKRDSDDEHHHHRHRSRKH
EDSDVEHGHHRHRSKRIKDEKTEVEETVSKSDQSLKASPGDNIPIYLQNEPSQVSD
ELRAKIRAMLADFLYSSN"
/note="gene_id:MC015.6
/known protein"
/known_start=1
/evidence-not_experimental
/protein_id="BAB08580.1"
/db_xref="GI:9758108"
/translation="MAHFMKLLVLTIALTTVVVIAIVTTTKTTATFALEDPFKDLTP
PQTIRPSRFLAERVDKGEKPKARNPAAHCHNDNEITCSDSYSTGANATMACNN
KCMDLSTDDCKGLCKNKCFQGTCCRGQCVVYVAYDKRHCGECNHRCDLGEPCVYGLC
NYA"
join(28332..28343,28421..28521,28615..28674,29082..29333,
29413..29598,29738..30006,30091..30426)
/note="gene_id:MC015.7
pir||T04808
strong similarity to unknown protein"
/codon_start=1
/evidence-not_experimental
/protein_id="BAB08581.1"
/db_xref="GI:9758109"
/translation="MLLKIKRPVTVSNYQKDETVEEGCGGRNCLSKCCINGARLPL
TKNLDKSVGENTESPVTFLSVIGEWEDREFGLFRYDVTACETKVIPIGKYGRIAQ
LNEGRHLKRPFRFVDKLOPFDGKNFTKVGQELLFQFKATNDSDSLQFLAS
WPLDANDSPSVAINVSPYEXHLLIRVLDCLPQRIDHRSLLALQMAEADNPYF
RYGNSLAFATINHLHQAYLHAMOFFPIEKASSIKITTTNNGVKISKLLNYPVRGLL
LEGDTIKDLADTVSDASVCQNNNIPFNILSDSKRIFLLPOCYAKAQLGEVSSST
LDTQNPVAVMMEGSHMVLKREDYEGASEEKAWRLLAEVLSSEERFREVTMIFDAI
GFSHEEEREEBQNSMNGSFTIVHCPVSKEEAVSN"
complement(join(32091..32480,32573..32689,32777..32895,
32993..33101,33200..33238,33373..33451,33563..33637,
33747..33811,33937..33992,34133..34225,34493..34745))
/note="gb|AAD18050.1
gene_id:MC015.8"
/codon_start=1
/evidence-not_experimental
/product="molybdopterin synthase sulphurylase"
/protein_id="BAB08582.1"
/db_xref="GI:9758110"
/translation="MMSNGDSSEIVRELEELKKAEIHRISTLEAKLQDTAAVEL
YDANGSYLTAPEVLEHGLSPDIYRSLPLSFVAGOSNLLKSSVLVIGAGGL
GSPALLYLAAACGOLGDIIDHVDVLLNNHROIITHEAFIGHPKVKSAARCSNST
IKVDEVEALRFSNALEILSQDIIVDADNPSPRYMSIDCCVLGKPLVSAALGME
GOLTVNHNNGPCYKCLFPTPTPTACQRCSDSGLVGVVPGICQALETIKLASL
AGPLNLLFALSRMIRIVIRGSOCTVCGDSSFNKOTFKDFYEDFTQPLF
AGPLNLLFAESRISSEKFEKILQKEQHVLLDVRPSHYKIVSLPDSLNIPLANLETR
LNELTSALKKGNHGANTSCNTPSVFVVCVRGNDQSRAVQVLRSEGSFSAKDIIGGL
EAWANVPNPEPT"
40087..40320
/note="contains similarity to 50S ribosomal protein L30
gene_id:MC015.9"
/codon_start=1
/evidence-not_experimental

```

alignment_scores:
 Quality: 89.00
 Ratio: 1.648
 Percent Similarity: 51.429
 alignment_block:

```

US-09-301-906-5 x AB010071 ..
Align seg 1/1 to: AB010071 from: 1 to: 82918

12 GlyLysLysValLeuMetGluAlaArgThrSerAlaThrSerAs 28
||||| ||||| ||||| ||||| |||||
7700 GGAACAAGGTT.....GCAGAAATTCTACATCTGG..... 7732

28 pValSerAspPheAspValValPheGluAlaValSerAsnAlaLeuVal 45
||||| ||||| ||||| ||||| |||||
7733 .....GATGCTGTCTCTCATGTG..... 7750

45 alValHisTyrHisArgValValProTyrAlaProValLysArgGluGln 61
||||| ||||| ||||| ||||| |||||
7751 .....GCACCTCAGAAAAGGCACCA 7771

62 ProLysProAlaValLysGlnAsp.....GluGlnLysProLysArgGln 76
||||| ||||| ||||| ||||| |||||
7772 GAGAACCTGCTCCCAAGCCTTCTCTCTCTGCTGAGAAAGCCCAAGTTGA 7821

76 nAlaSerHisTrpAlaValLysProThrAlaValGlyValHisValProL 93
||||| ||||| ||||| ||||| |||||
7822 AAGTACCAAGTTTGCAGAGAAACCAAGGCACCATCACCGCCACCTCCAC 7871

93 euProLysLysGlnGluAlaLeuGluProAlaGlnSerValProGlnGln 109
||||| ||||| ||||| ||||| |||||
7872 CGCATCCAAACAGTCAGCTAAGAACCG.....CAGCTTCCACCTAAG 7915

110 SerLeuGluGluLys 114
||| ||||| ||||| ||||| |||||
7916 GATAGGAAGACCG 7930

seq_name: gb_p12:ART223803

seq documentation block:
LOCUS ART223803 1724 bp mRNA PLN 28-JAN-1999
DEFINITION Arabidopsis thaliana mRNA for 2-oxoglutarate dehydrogenase E2
subunit.
ACCESSION AJ223803
VERSION AJ223803.1 GI:4210331
KEYWORDS 2-oxoglutarate dehydrogenase; E2 subunit.
SOURCE thale cress. thaliana
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
Rosidae; Capparales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 1724)
AUTHORS Machuy N.
DIRECT SUBMISSION
TITLE Submitted (28-JAN-1998) Machuy N., Max-Planck-Institut fuer
JOURNAL Molekulare Pflanzenphysiologie, Karl-Liebknecht-Str. 24-25/Haus 20,
14476 Golm, FRG
REFERENCE 2 (bases 1 to 1724)
AUTHORS Machuy N., Klein M. and Mueller-Roeber, B.
TITLE Cloning and characterization of 2-Oxoglutarate dehydrogenase from
Arabidopsis thaliana
JOURNAL Unpublished
FEATURES
source
1..1724
/organism="Arabidopsis thaliana"
/cultivar="C24"
/db_xref="taxon:3702"
/cell_line="558"
/cell_type="enriched epidermal fragments"
/clone="2-OGDH E2"
/clone_lib="Ara558-EF-Zap"
90..1478
/EC_number="2.3.1.61"
/codon_start=1
/product="2-oxoglutarate dehydrogenase E2 subunit"
/protein_id="CA11553.1"
/db_xref="GI:4210332"
/translation="MWLRVFRASIRGSSASGLKSLQSSRLVASSQFHSVATET
CDS

```



```

seq_documentation_block:
LOCUS      AE001935      10165 bp      DNA      BCT      22-NOV-1999
DEFINITION Deinococcus radiodurans R1 section 72 of 229 of the complete
            chromosome 1.
ACCESSION  AE001935 AE000513
VERSION    AE001935.1 GI:6458517
KEYWORDS
SOURCE
ORGANISM   Deinococcus radiodurans.
REFERENCE  1 (bases 1 to 10165)
AUTHORS    White,O., Eisen,J.A., Heidelberg,J.F., Hickey,E.K., Peterson,J.D.,
            Dodson,R.J., Haft,D.H., Gwinn,M.L., Nelson,W.C., Richardson,D.L.,
            Moffat,K.S., Qin,H., Jiang,L., Pamphile,W., Crosby,M., Shen,M.,
            Vamathevan,J.J., Lam,P., McDonald,L., Utterback,T., Zalewski,C.,
            Makarova,K.S., Aravind,L., Daly,M.J., Fraser,C.M. et.al.
TITLE      Genome sequence of the radioresistant bacterium Deinococcus
            radiodurans R1
JOURNAL    Science 286 (5444), 1571-1577 (1999)
MEDLINE    20036896
REFERENCE  2 (bases 1 to 10165)
AUTHORS    White,O., Eisen,J.A., Heidelberg,J.F., Hickey,E.K., Peterson,J.D.,
            Dodson,R.J., Haft,D.H., Gwinn,M.L., Nelson,W.C., Richardson,D.L.,
            Moffat,K.S., Qin,H., Jiang,L., Pamphile,W., Crosby,M., Shen,M.,
            Vamathevan,J.J., Lam,P., McDonald,L., Utterback,T., Zalewski,C.,
            Makarova,K.S., Aravind,L., Daly,M.J., Minton,K.W.,
            Fleischmann,R.D., Ketchum,K.A., Nelson,K.E., Salzberg,S.,
            Smith,H.O., Venter,J.C. and Fraser,C.M.
TITLE      Direct Submission
JOURNAL    Submitted (08-NOV-1999) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES   Location/Qualifiers
            source          1..10165
                        /organism="Deinococcus radiodurans"
                        /strain="R1"
                        /db_xref="taxon:1299"
                        /chromosome="1"
            gene            complement(81..323)
                        /gene="DR0803"
            CDS              complement(81..323)
                        /note="identified by Glimmer2; putative"
                        /codon_start=1
                        /transl_table=11
                        /product="hypothetical protein"
                        /protein_id="AAFL10390.1"
                        /db_xref="GI:6458526"
                        /translation="MSVTRQELLALALGQLSDDQQRVQALDHPALRTALRRDLDA
                        LTSLLADLDVQAQVPPGADCGSCGPKPPSSR"
            gene            complement(320..1168)
                        /gene="DR0804"
            CDS              complement(320..1168)
                        /note="similar to GB:AL123456 percent identity: 51.81;
                        identified by sequence similarity; putative"
                        /codon_start=1
                        /transl_table=11
                        /product="sigma factor, putative"
                        /protein_id="AAFL10382.1"
                        /db_xref="GI:6458518"
                        /translation="MSADPPSHAPPAGPFTSDGARPGPAPFPSPDSQVLRDWAAGHE
                        GALLEHRYRALLGYLGGRIILASPEAIQQGVDDALMTAWHAAHYEPARASVHGWL
                        STAHHFLRGLRGAGGTGLRHQFGRDTPGDQDHLALAYYCGPPPARLARIISGLP
                        QPAVEEALLSATLSRLPGLAQEALAKAAPAPAPPAEPTAEAEEDDLFPTLSSLAATAP
                        VTRVEWPPRRPEQREPEAEAVLPEASEATLPVSEPLPNAEPATQAEQSPETT
                        APEVQA
            gene            1336..1734
                        /gene="DR0805"
            CDS              1336..1734
                        /gene="DR0805"
                        /note="similar to GB:AE000511 PID:2313428 percent
                        identity: 83.33; identified by sequence similarity;
                        putative"

```

```

/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAFL10383.1"
/db_xref="GI:6458519"
/translation="MYCGPSRFRPCSRFRSHFWRNMSLSGPFPIILLVVIALL
FGAKILPELMKMGQIGREFKNEVKDPAAPVPLQGPQLQPGQTAGTTVTAQPVVDV
ASROLDPMIDPVTGAPVNVPEVQVPGERR"
1775..2587
/gene="DR0806"
1775..2587
/note="DR0806"
PID:148237 GB:U00096 percent identity: 56.56; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="Sec-independent protein translocase protein
TatC, putative"
/protein_id="AAFL10384.1"
/db_xref="GI:6458520"
/translation="MTQLPPPEQTVLKPAPPELASAPLFDHLEELRRRLLSVVFLAV
GMVATYRVQLIELVLYPLTYSELITTKRVQLVTKLASQLLSFNLAFWAGLTAL
PFTVQWIAFIAPGLYPOERRRGLPFLTGAGFAFAAGVFGYKLVLPMTVPFTIEFLA
GTVTQMDLOEYIGTVYFVAFVAFELPILAVILTRLGIVNHTMLRQGRFALIGI
MILAAVITPTDPANMALVAVPLYALVELGVLSRVFRTIAPEEQERPAPMS"
2633..3652
/gene="DR0807"
2633..3652
/note="DR0807"
/note="similar to PID:1651935 percent identity: 56.25;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="prolipoprotein diacylglycerol transferase"
/protein_id="AAFL10385.1"
/db_xref="GI:6458521"
/translation="MPRRGPFALFAGCVSRGASRSPGHTLTRRPMQVFLQIGNFT
IAWGVLMGLGIAAGYVGLKLAERGLNADLPMTLWMLFWGFGVGLRVFLTSWH
IFEGIPFPRVLLDIYNLRNGGSIHGGLIGVLTLYFARRYRLNFYQADLAVPGVA
FGIIGRLGNIMGTDVTRVGTWPGIYHWPASARAFHEGMCRRPNPMDLSKYCHE
VGGQIMTAPVHTQYLVIGITILAVASYFLKSRVPGWAFQFWLWYSLIRAGWEE
TFRNLPLTIKTYLNOGLDAPGIGLWTDQTQIFSVPLILVLSLWMLRRLRQRQVTVPTVP
VQSQS"
3782..5245
/gene="DR0808"
3782..5245
/note="DR0808"
PID:467439 percent identity: 62.28; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="UDP-N-acetylglucosamine pyrophosphorylase"
/protein_id="AAFL10386.1"
/db_xref="GI:6458522"
/translation="MTMTQTRPLDVVILAAGQTRMNSALPKVLHPVAGRPMAWA
VKAARDLGARKIVVVTGHAQVEAVVQPGVAFARQERGLGTGDAFLSGASALTEGG
ADILVLYGDTPLLRVPTDLRELVAHQSGAMTVLGTQELPDATGIRVIRGEGGAVER
IVQKQASDAEKAIGEFGNVYFDERAPELARRIGNDNKAGEYLYLTDLGLYRAGGA
TVRAFLLSDPDEVILGANDRAGLTQLEATLQRRINEGHLKAGVSTSLPGTVLIEDTVEI
GRDVTPEPGLAGQTRIAGCAVIGAVSVITDSVTHFRAVIKAHVLEQAEVAGSDV
GPAPRLRQSVLGEVGHGIVNFVKNARLDAGVKAGHLAVLGDVEIGAEINVCAGTIV
ANFDGLNKQSKVAGVFGSNTTTLTPRVVGDAAFIAGGSAIHDDVPEGAMAVARGK
QRNIEGWARIYWAAPERREKRLAKKLPWLAGWLSRTQG"
5274..6272
/gene="DR0809"
5274..6272
/note="DR0809"
/note="similar to GB:AE000666 percent identity: 50.23;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11

```



```
/product="conserved hypothetical protein"
/db_xref="AAFI0387.1"
/translation="MEGRRGQSGRPFSIAVVTLLPCPGHLPAAVRLQCPRMASAPP
RLILPLSLVTSRPAWNTIGLTVGVLTRGLTLDGGVLAALLLTLTPENLILY
GLNDLAAREEDARRSKGQCARLPPEEGPLRSTLLINAPRLILGLLAWLPFRATA
TLLLAALFVGLSPRLPKARFPLDGLSNAYALPALPALAGAPFPWAPLLALMS
YSVGRHAFDAQDIIADRAAGTRVATTIGTGVAAYALAWFTLAGALLVPVSRUTAL
ALWITCGGMALALLRRPAEQAGRLPLYSIVTPWIVGVSAGVQLVYLLARGQWDL"
6257..7624
/gene="DR0810"
CDS
6257..7624
/gene="DR0810"
/notes="similar to GB:221955 PID:288223 percent identity:
51.52; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="dehydrogenase, putative"
/db_xref="AAFI0388.1"
/translation="MGRPVKSPVESVILGGGLAGLALACLILAGRHAVTVYERDRAG
GKLRRAEAGGLTFDTPSLTFPEVWRALLARIGEDPLELRPLPGGLGVQVTPFGAL
PLVPPEHPPLACEMORYVAVSPRLSHLLALTTPRLTSPDFLRASRALAGVLGPHL
SAESNTAANDFAALAHARTOALNAGLPQDAPALYALLGLIGAEVYRPARGWAV
LDALLDFARRGVTVREGTEVYRLSGQITLSGGEVCRHDLVSLDPLARLARQGS
DASPLARRTVSGLALFALNAGESGLPATSVPDGFVFRSLHAGTLPDPTLTILVHA
EGKLAVLTVFALSHLSHPWQAOQLRRVETVLHAPGLIQGAATLADPLHVAAG
GHPGGAIFYAAAPLWRAGFPHPQYRLAPLQWGVGTGVHPGGGLPAIFGGLIVDGLL
ALIRL"
complement(7679..8023)
/gene="DR0811"
CDS
complement(7679..8023)
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/db_xref="AAFI0391.1"
/translation="MSPVLRTALLTLNSVWMTFAWYGLLGKMAARPLLAVLASWGI
AFPEYLQVPAANRIGHQVLTVAQLKMQVEISLAVFVFPFVLYLREPLRLNYLWALC
MNGAVYVFRGR"
complement(8185..9960)
/gene="DR0812"
CDS
complement(8185..9960)
/notes="similar to GB:K02496 SP:P00782 PID:142526
PID:773560 percent identity: 59.94; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="serine protease, subtilase family"
/db_xref="AAFI0389.1"
/translation="MKRTKTLTALLTLALSCGQQLHTPPTAGQATSPQCADNEV
alignment_scores:
Quality: 83.50 Length: 124
Ratio: 1.176 Gaps: 7
Percent Similarity: 57.258 Percent Identity: 26.613
alignment_block:
US-09-301-906-5 x AE001935/rev ..
Align seg 1/1 to reverse of: AE001935 from: 1 to: 10165
4 SerValSerAlaLeuArgGlyAspGlyLysLysValLeuMetGluAlaAr 20
:::||||| ||| ||| ::::::::::::::| |::|
658 GCCATCAGCCGCTGCCGGGACTCGCACAGGAGCGTGGCAAGCGGC 609
20 gThrSerThrSerAla.....ThrSerAspValSerAsp 32
::: ::| ||| |||:::|:::|:::|:::|
608 AGTGCCTGCCCGCCCGCCGCGCAGAACCCAGCGGAGCGCAGAG. 560
```

```
32 heAspValValPheGluAlaValSerAsnAlaLeuLeuValValHisTyr 48
||| ::| ||| ::|::|::|::|::|
559 ..GAGGACCTCTCCCAACCTGTCACG..... 533
49 HisArgValValProTyrAlaProValLysArgGluGlnPro...LysPr 64
:::|::|::|::|::|::|::|::|::|::|::|::|::|
532 .....CTCGCGCCACCGCGCGGTACGCGGGTGGAAACCTGGGAACC 489
64 oAlaValLysGlnAspGluGlnLysProLysArgGlnAlaSerHisTyr 81
|::: ::|::|::|::|::|::|::|::|::|::|::|::|
488 GCCCGCGCGCCGGAACAGCGGAGCGGAGACAGAGCGGAG.....G 445
81 laValLysProThraL.....ValGlyValHisValProLeu 93
|||::|::|::|::|::|::|::|::|::|::|::|::|
444 CGGTGCTCGCGGAAGCTTCCGAGCGCACTTTGCTGTGTTTCAGAACCTCTC 395
94 ProLysLysGln.....GluAlaLeuGluProAlaGlnSerValProGlnG 109
|||::|::|::|::|::|::|::|::|::|::|::|::|
394 CCTAACGCGAAGCCGGAAGCACCACCCGCGGAGCAAGCCGCGGAGAC 345
109 nSerLeuGluGluLysAlaLa 116
::: |||
344 CACCGCGCGCGAGGTCCAGGCA 323
seq_name: gb_htg3:AC010847
seq_documentation_block:
LOCUS AC010847 173807 bp DNA HTG 13-DEC-1999
DEFINITION Drosophila melanogaster chromosome X clone BACR10M08 (D904) RPCI-98
10.M.8 map 18D-18B strain Y; cn bw sp, *** SEQUENCING IN PROGRESS
***, 101 unordered pieces.
ACCESSION AC010847
VERSION AC10847.10 GI:6563418
KEYWORDS HTG; HTGS_PHASE1.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 173807)
AUTHORS Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,
Richards,S., Seth,H., Swirskas,R.R., Wan,K.H., Webster,D.,
Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
Sequencing of Drosophila melanogaster
Unpublished
2 (bases 1 to 173807)
AUTHORS Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequeira,A., Seth,H., Snir,E.,
Swirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.
Direct Submission
Submitted (24-SEP-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Dec 13, 1999 this sequence version replaced gi:6503258.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive web site (http://www.fruitfly.org/sequence/) or send email
to bdg@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 200 bases.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 101 contigs. The true order of the pieces
* is not known and their order in this sequence record is
```

* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
* 497: contig of 497 bp in length
* 498 577: gap of unknown length
* 578 1222: contig of 645 bp in length
* 1223 1302: gap of unknown length
* 1303 1391: contig of 689 bp in length
* 1392 2071: gap of unknown length
* 2072 2644: contig of 573 bp in length
* 2645 2724: gap of unknown length
* 2725 3571: contig of 847 bp in length
* 3572 3651: gap of unknown length
* 3652 4118: contig of 467 bp in length
* 4119 4198: gap of unknown length
* 4199 4708: contig of 510 bp in length
* 4709 4788: gap of unknown length
* 4789 5429: contig of 641 bp in length
* 5430 5509: gap of unknown length
* 5510 6178: contig of 669 bp in length
* 6179 6258: gap of unknown length
* 6259 6989: contig of 731 bp in length
* 6990 7069: gap of unknown length
* 7070 8310: contig of 1241 bp in length
* 8311 8390: gap of unknown length
* 8391 9264: contig of 874 bp in length
* 9265 9344: gap of unknown length
* 9345 9845: contig of 501 bp in length
* 9846 9925: gap of unknown length
* 9926 11279: contig of 1354 bp in length
* 11280 11359: gap of unknown length
* 11360 12063: contig of 704 bp in length
* 12064 12143: gap of unknown length
* 12144 12888: contig of 745 bp in length
* 12889 12968: gap of unknown length
* 12969 13997: contig of 1029 bp in length
* 13998 14077: gap of unknown length
* 14078 14660: contig of 583 bp in length
* 14661 14740: gap of unknown length
* 14741 16172: contig of 1432 bp in length
* 16173 16252: gap of unknown length
* 16253 17216: contig of 964 bp in length
* 17217 17296: gap of unknown length
* 17297 18375: contig of 1079 bp in length
* 18376 18455: gap of unknown length
* 18456 19931: contig of 1476 bp in length
* 19932 20011: gap of unknown length
* 20012 21401: contig of 1390 bp in length
* 21402 21481: gap of unknown length
* 21482 22584: contig of 1103 bp in length
* 22585 22664: gap of unknown length
* 22665 24179: contig of 1515 bp in length
* 24180 24259: gap of unknown length
* 24260 25185: contig of 926 bp in length
* 25186 25265: gap of unknown length
* 25266 26305: contig of 1040 bp in length
* 26306 26385: gap of unknown length
* 26386 27257: contig of 872 bp in length
* 27258 27337: gap of unknown length
* 27338 27949: contig of 612 bp in length
* 27950 28029: gap of unknown length
* 28030 29026: contig of 997 bp in length
* 29027 29106: gap of unknown length
* 29107 30504: contig of 1398 bp in length
* 30505 30584: gap of unknown length
* 30585 31791: contig of 1207 bp in length
* 31792 31871: gap of unknown length
* 31872 32803: contig of 932 bp in length
* 32804 32883: gap of unknown length
* 32884 34157: contig of 1274 bp in length
* 34158 34237: gap of unknown length
34238
* 35253: contig of 1016 bp in length
35254
* 35333: gap of unknown length
35334
* 36894: contig of 1561 bp in length
36895
* 36974: gap of unknown length
36975
* 38847: contig of 1873 bp in length
38848
* 38927: gap of unknown length
38928
* 40303: contig of 1376 bp in length
40304
* 40384: gap of unknown length
40385
* 42460: contig of 2077 bp in length
42461
* 42540: gap of unknown length
42541
* 45210: contig of 2670 bp in length
45211
* 45290: gap of unknown length
45291
* 46703: contig of 1413 bp in length
46704
* 46783: gap of unknown length
46784
* 48483: contig of 1700 bp in length
48484
* 48563: gap of unknown length
48564
* 49341: contig of 978 bp in length
49342
* 49621: gap of unknown length
49622
* 51424: contig of 1803 bp in length
51425
* 51504: gap of unknown length
51505
* 54181: contig of 2677 bp in length
54182
* 54261: gap of unknown length
54262
* 57237: contig of 2976 bp in length
57238
* 57317: gap of unknown length
57318
* 59095: contig of 1778 bp in length
59096
* 59175: gap of unknown length
59176
* 62804: contig of 3629 bp in length
62805
* 62884: gap of unknown length
62885
* 65116: contig of 2232 bp in length
65117
* 67196: gap of unknown length
67197
* 67276: gap of unknown length
67277
* 69681: contig of 2405 bp in length
69682
* 69761: gap of unknown length
69762
* 71889: contig of 2128 bp in length
71890
* 71969: gap of unknown length
71970
* 73832: contig of 1863 bp in length
73833
* 73912: gap of unknown length
73913
* 76588: contig of 2676 bp in length
76589
* 76668: gap of unknown length
76669
* 80087: contig of 3419 bp in length
80088
* 80167: gap of unknown length
80168
* 83703: contig of 3536 bp in length
83704
* 83783: gap of unknown length
83784
* 87270: contig of 3487 bp in length
87271
* 87350: gap of unknown length
87351
* 90838: contig of 3488 bp in length
90839
* 90918: gap of unknown length
90919
* 96597: contig of 5679 bp in length
96598
* 96677: gap of unknown length
96678
* 102202: contig of 5525 bp in length
102203
* 102282: gap of unknown length
102283
* 106529: contig of 4247 bp in length
106530
* 106609: gap of unknown length
106610
* 112127: contig of 5518 bp in length
112128
* 112207: gap of unknown length
112208
* 119794: contig of 7587 bp in length
119795
* 119874: gap of unknown length
119875
* 128971: contig of 9097 bp in length
128972
* 129051: gap of unknown length
129052
* 138913: contig of 9862 bp in length
138914
* 138993: gap of unknown length
138994
* 150832: contig of 11839 bp in length
150833
* 150912: gap of unknown length
150913
* 151242: contig of 330 bp in length
151243
* 151322: gap of unknown length
151323
* 151833: contig of 511 bp in length
151834
* 151913: gap of unknown length
151914
* 152228: contig of 315 bp in length
152229
* 152308: gap of unknown length
152309
* 152758: contig of 450 bp in length
152759
* 152838: gap of unknown length
152839
* 153233: contig of 395 bp in length

anthelmintic macrolide avermectin in Streptomyces avermitilis
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9509-9514 (1999)
99380548
2 (bases 1 to 64957)
Ikeda, H., Nonomiya, T., Usami, M., Ohta, T. and Omura, S.
Direct Submission
Submitted (13-SEP-1999) to the DDBJ/EMBL/GenBank databases. Haruo
Ikeda, School of Pharmaceutical Sciences, Kitasato University,
Microbial chemistry; 5-9-1 Shirokane, Minato-ku, Tokyo 108-8641,
Japan (E-mail: ikeda@mc.pharm.kitasato-u.ac.jp, Tel: +81-3-5791-6242,
Fax: +81-3-3444-6197)

FEATURES
source
1..64957
/organism="Streptomyces avermitilis"
/db_xref="taxon:33903"
101..12019
/gene="aveA1"
101..12019
/gene="aveA1"
/note="contains a loading module, Ave module 1 and module
2; multifunctional polyketide synthase"
/codon_start=1
/transl_table=11
/product="type I polyketide synthase AVES 1"
/protein_id="BA84474.1"
/db_xref="GI:5902891"

/translation="MORMDGERPRPAAGVILGVADEADGGVVFVFPGGQPMGMR
ELLADSGVRESVRACEAFAPYVDMVSEVQLRSDPADGLDRVDVQPTLFAVMSLS
AALMRSGQVEPCAVIGELIAAAHVSGQLSADAAVVTWLSQAOVTTLAGTALVS
VAATPDELPRTAPMTEDNPARLAAVNGPRSTVYSGAREAVADLVADLTAAQVTR
MIPVDPAISPLMAIEERVVSGLLIPITPRSPRIPTHSSVYTGRLDTRREDAAYWYTR
MSTVFEPAARLLQGGPKTFVEMSPHVLTMGLQELAPVLTGTADTVIMGTNR
RGRTDHLFTLSAQRHGETSATTVL SARLTALSPQOQSLLDLVRAHMAVLND
DGNETASDAGSASAHILGDFSMGVELNRKATGLRPLVLLFDHTTAAVAAR
LRTAALHDEDTAPYDPSGHGTAAADDPITAIIGMACRPGVRSKDLWELAAAS
GDAIGCFPTDQMTQORHAODPTQPTFYPOGGFLHDAHFDAGFCGISPREALA
MPOQRLLLETWEAREAGIDPLSVRGSRGTGFAGALSFDYGPMDTASSBGAADVE
GHLLTQVLSGRJATISFGLGEPAITVDITGCSALSTLHLACQSLRSGCTLALAG
VSMSTLCMFTEFSRQSLGSDGRCAY SAAADGTGWEGVGMLLVERLSDAVRIGH
RVLAUVAGANQDASNGLTAPNGPAQERVTIQRALANAGSVAADVVEGHGTGTL
GDI IEAQUALTYGQAGDRPLWGLSKNIGHTMAAGGVGIKVMALRECVLPRT
LHWDFSPQVDSAGAVRLLTEAVPMPGDAAGRLRAGVSFGIGCTNAHVILLEEAPA
AGCVAGGVLEGACGLAISVAESVAAPVAVSNPVASVPPVPPVPPVPSARSAG
LRAQALQYIVAVRDPVSLADVAGLACRAVLEHRAVVLAAADREELVQGLALAG
EPRVYTHGAPGDRGVVFPFGQGMAGVRLLASPPFARMQACEALAPW
VDMVYDILDRDAGDAWERADYVQVLFVSMVSLAALWRSYCIEDPAVLGHSQGEIA
AAHCCALSKDAAKTVLRALAAVRGGMASVPLPAQEVHQLIGRWAGRIWA
AVNGPSTAVSGDAEAVLAYCAGTVRARI PVDYASHCPHQPPLREELLELGD
ISFPQSGVFESTVEGTWDTTLDAAWYRNHLQVRFSDVAQALADGHRVFEVS
PHTLVPALEDITTEADVTAIGSLRRGNDTRRELTALAHHTTGTGTPTWHHYH
THHHHPHTHLDLPTYPFQHQYWLSESPQSGAGSGAGSGAGSGAGTAGTGTGA
EVESREWDAAVARDLETVATLAVPPSAGLDTVPALSAWHRHQDQARINTWYQET
WKPLTPTTHQPHOTWLI AIPETQTHPHITNLNLHHGHTPIPLTNHTHTNQH
LHHTLHTRQQAQNHHTGAILGLSLLALDETPHHPHTPTGTLNLLILTOHTHTH
PTPLWYATNATHTHPNDPLHTQATQWGLARTLLEHHPHTHAGIIDLPTPTPTPT
LOHLTLPQHQHOLAIRTTGTRHRLTPTTLTHQPTPTPTPTPTPTPTPTPTPTPTPT
ANTLHHTLTHQPTQLHLSRTSGPTPHQAHLTLOQKGIHLTITCDTSPNDQLQ
QLNLTTPQPHLTPTVHTAGILDDATLTNLPTQLNNVRARAKASHLHQLTQHTPL
TAFVLSAAATGAPQOYAAANAYLDLAHHRHTHLPATSIANGTWQGLADS
DKARVSLDRGFRPMSPELATAVQAIADTERPVYV IADIDWSKIEHTSQSDLYSA
ANEREPVORPTPPELHHTLAHQTSADQALRELVRAVLAADHPKAPADQDS
FRALGDSUTAVEPILKATGLRPLVSLVFDHPHTPAKLAHLQGLRGTAAESAPL
AAVATAEASVTEPIATVGMACRPGCVTSADDFWDLISSEODAIQQFTDWDLDTL
YDPPDHQPTGTRNGGFLYDAGHDAEFGISPREALAMPQOORLLLETAWEITIEHA
GNPHPTLHGTPTGVFTNGQDYALRVHAGOSTDGFALTGTAGSVISGRISYTFGFE
GRVAVSDTACSSLVALHACQALRAGESCMALAGVTVNPSPGAFVESRQRLAAD
GHCKAFSAADGTGWEGVGMLLVERLSDAHRNHRVLAUVAGSNVQDAGSNLTP
NGPSQQRVIRQALANAGLSAGDVAEAEHGTGTT LGDP IEAQUALTQDRAEGEPL
WLGSVKSNYCHTQAAAGVAGIKVMNLRHGLLPRTLHVDESPSHVDWAGAVQLLE
TVPWPGCEGRLLRAGVSFGCTNAHVILEEAPADVPDGPAGGAGSDDEAAAG
SPGVPMPLVSAKSOPALRAQALHAHLTDHPCLDLADVGYTLAHARAVFDRATLIA
ARDTFLQALQALAGEPHPAVHSSAPGCTGGEAGKTAFICSCQGTQRGMHGL
YHTHPFAALNDICTHLDHLDHPLLLPTLQNDNDNDEAALLLQOTRYQALFAFQ

VALHRLTDCGYHTPHYAGHSLGEITAAHLAGILTLDTATTLITQATLMTQMPPTGT
MTLHTPHHTHHLTAHENDLAIANINTPSLISGTHTVQHIHTLTCOOQIKTKT
LPTNHAHSPHTNPILNQLHQHTQTLTYHPHTPLTANTPPDQLLTLPWYTOQARNT
VDYATQTLHOHVTTYIELGPDNTLTTLTHNLNPNPTTLTLTPHHPHTQHLTT
NLAQTTTTWHPHYTHDNOPTHHTLDLPTYFQHHHVWLESTQPCAGNHAAGLDP
TEHPLGATLELATDGCALLAGRLSLRSHPLADHAVGTVLLSGATFLELALAGTY
VGCDRDEULTLHAPLVVPPDGVSVQVGAADGEGRRLVSVYARGSGAGGGGAGG
WTCCHASGLVLEAAGVVVDGLAGVMPPRGAVADVGVDRDLGACGLVGFVFSGL
RAVRDGGDLLEAVCLPEAGWDAAGFGLHPLLDGVVQPLSVLLPGGTGFGSAGFL
EGVRVAPWGVGSLHRAGTVGVRRVSVAGRGREAVSVVVGDEAGVPVAVDRLEL
RPVDMQLRAVSVSAGRGSIXAVOMAEVGPVPCGOAMAHEDVCGSGGPPGVVV
LRCPDVAGAGGGGGGGGVEGVGVGLVGVQVWGLERFAGSRLVVVTRGAVVAGE
DGPVDVGSVGLVRSQAQHPDFRVLDDTDTGTDGAGAGVGDGKRAVAV
ACGEPQLAVRGERLLAARLKLESSGDVPAQRSGDTRARSDDPAPRSGPPARSVD
VSGREVLPMWLSGGSVLTGTVGLAAVRAHLAGVCGVRLDILLVSRGPDAPAGELR
AELAAALGAEVRIACDVGRERREVRLLLEGVAPCGPLTVGVVHAAGDADTASITPER
LCTVFAAKVDAALLDLDELTHGMELSAFVLFSSAAGILGSAGOCNYAANAALDALAYR
RRAAGLPVSLANGLWEASGWTGHLAGTDHRRRIIRSGLHPMSTPDALALFDAALD
RPVLLPADURPAPPLPPLLODLLPATRRRTTRTTTGGADNGAGLHARLAGOTHEQH
TLLALVRSHIATVLTGHTTPTPPDRAFDLGDSLTAVELNRNLRTGLRPTTL
AFDPNPTTLTHRLTQLOPQDNVAPVLAELDKLESALSALDKTDSASERVTLRLK
SLMRWNAPOHPTAESADDDDEKFTSATEAEIFKFDIDLGLS"
12071..30790
/gene="aveA2"
12071..30790
/gene="aveA2"
/note="contains Ave module 3, module 4, module 5 and
module 6; multifunctional polyketide synthase"
/codon_start=1
/transl_table=11
/product="type I polyketide synthase AVES 2"
/protein_id="BA84475.1"
/db_xref="GI:5902892"
/translation="MOLANEKLELYKRVADLDRRLRYEVVEREQEPIAIVGMA
CRYPGATSPTRLHLVKSDTAIGFPTDRGNLRLQVDPDORSGTSYTRSGGLY
DAGDFAAEFELSPREALAMPQOORLLLETTWTFEGQGDIDPRMSGRSGVLRGPNP
EDITTYTHQPSNAVEGYLLDTSAASTIASGRISYFNFGLEPAITIDTACSSVLHL
ACQALRSGCTMALAGGASVMAATPFVTFESRQGLAANDGRKAFSAAADGTGMSGV
GMLLVERLSDAHRHRLVAVRGSVNDGASNGLTAPNGRQSVIRQALANAHLS
PADVDAEARGHTGLTLDPIEAQALVEAYGDRPNGRPLWGLTKNIGHNMAAGVSG
GLVKVMALNRGLTLPRLHVDPEPSHVDMSAGVOLLTETVPMWPGEGRLRRAAGVSF
GVSGTNHVILEEAPAHNTPSDTPADDVFGESAADEADGSGDEAAGSPGVWPLVSA
KSPALRAQALHAHLTDHPLDGLADVGYTLAHARVDFHRLTAAADRTFLQALQ
ALAAQEPHFAVHSSAPGCTGGEAGKTAFICSGQGTQRGMHGLYHTHPFAAL
NDICTHLDHPLHPLLTQDPTQDTTLEEAALLQOTRYQALFAVQALHRL
LTDGYHTPHYAGHSLGEITAAHLAGILTLDTATTLITQATLMTQMPPTWTLHT
TPHTIHTLTAHENDLAIANINTPSLISGTHTVQHIHTLTCOOQIKTKLTPTHA
FISPHQNPILNQLHQHTQTLTYHPHTPLTANTPPDQLLTPTHTQQAQNTVDYAT
TQTLHQGVTTYIELGPDNTLTTLTHNLNPNPTTLTLTHHPHTQHLTLTLAKTT
TWHPHYTHHHNQHTHLDLPTYFQHHHVWLEPQKPSDPSSEGEQATPTST
PLRDVLVGSQPODEELLRLVTHAAVGLHATPEVIVPNKAFKELGFDLSAAIOLR
NRLAADVLPPLPATLIFDYPMALCQFLRAATVGDATGTTTLPLTAVPADRPIATV
PLRDVLVGSQPODEELLRLVTHAAVGLHATPEVIVPNKAFKELGFDLSAAIOLR
GNACRYGDOVRTVDLQWVSGCHDAIGGFTPNRGHDLDTLVNPDPHGTSYTRSG
FLYDAGNPDDEFGISPREALAMPQOORLLLETTWTFESHACINPSLGRTPGVFAG
LTHYDAARFPTAPAGFEGYLGHSAGSASVAYALGLEPALVDTACSSVLAL
HLACQALRSGECSMALAGVTVNTPAGVFEFSRQGLAANDGRKAFSAAADGTGGE
GMLLVERLSDAHRHRLVAVRGSVNDGASNGLTAPNGRQSVIRQALANAHLS
LTPADVDAEARGHTGLTLDPIEAQALVEAYGDRPNGRPLWGLTKNIGHNMAAGV
GVGIKVMALNRGLTLPRLHVDPEPSHVDMSAGVOLLTETVPMWPGEGRLRAGVS
SGVSIKVMALNRGLTLPRLHVDPEPSHVDMSAGVOLLTETVPMWPGEGRLRAGVS
SGVSGVPMPLVSAKSOPALRAQALHAHLTDHPLDGLADVGYTLAHARVDFHRLT
IAADRTFLQALQALAGEPHPAVHSSAPGCTGGEAGKTAFICSGQGTQRGMHGL
YHTHPFAALNDICTHLDHLDHPLLLPTQDPTQDTTLEEAALLQOTRYQALFAV
PALFAFOVALHRLTQDGYHTPHYAGHSLGEITAAHLAGILTLDTATTLITQATLMT
QMPPTWTLHTTPHTIHTLTAHENDLAIANINTPSLISGTHTVQHIHTLTCOO
QGKTKTLFTKNAFHSPHNPILNQLHQHTQTLTYHPHTPLTANTPPDQLLTPTHT
TQQAQNTVDYATQTLHQGVTTYIELGPDNTLTTLTHNLNPNPTTLTLTHHPHH
PQTLHTLNAKTTTTHHPHYTHHHNQHTHLDLPTYFQHHHVWLESTQPCAGSG
SSGSGRAGTAGTAESRFDVAVARQDLTEVATTLAVPPSAGLHTNTPALSAWHRH
QDQARINTWYQETWKPLTLPHTHOPHOTWLI AIPETQTHPHITNLNLHHGHT
PIPLTNHTHTNQHHLHTRQQAQNHHTTGTGTLNLLALDETPHHPHTPTGTLN
LIDTQTHTPTPTLWYATNATHTHPNDPLHTQATQWGLARTLTLTPHTQPPPTPHG
ILDPTTPTHTLHLLTQTLQPHHQTQALRTTGTTRRLTTLTTLTPHTQPPPTPHG


```

/standard_name="D6S2267"
/note="sy24b3-L:The location is between each flanking
site of PCR primers."
/db_xref="GDB:5583975"
STS
57335. .57660
/standard_name="D6S1944"
/note="WI-7404:The location is between each flanking site
of PCR primers."
/db_xref="GDB:679168"
STS
57336. .57491
/note="RH17987:The location is between each flanking site
of PCR primers."
/db_xref="GDB:4572735"
complement(81490..81630)
STS
/note="SHG-17229:The location is between each flanking
site of PCR primers."
/db_xref="GDB:1231045"
complement(88508..88638)
STS
/note="SGC34813:The location is between each flanking site
of PCR primers."
/db_xref="GDB:4584664"
BASE COUNT 28701 a 22059 c 22429 g 26811 t
ORIGIN

alignment_scores:
  Quality: 82.50 Length: 103
  Ratio: 1.557 Gaps: 4
  Percent Similarity: 51.456 Percent Identity: 25.243

alignment_block:
US-09-301-906-5 x AP000517/rev ..
Align seg 1/1 to reverse of: AP000517 from: 1 to: 100000
15 ValLeuMetGluAlaArgThrSerAlaThr..... 26
: : : : : : : : : : : : : : : : : : : : : : : :
77481 CTTATCCTCACAGCAGGCTTCCTCCATCTCTTGGCAGACTTTCAGTC 77432
27 .....SerAspValSerAspPheAspValValPheGluA 38
: : : : : : : : : : : : : : : : : : : : : : : :
77431 TCCATGGCATTTCTCTCCATTCCTACCCCTTCCAGTCACCTTCCAAT 77382
38 laValSerAsnAlaLeu.....ValValHisTyrHisArg 50
: : : : : : : : : : : : : : : : : : : : : : : :
77381 CTGCTGGGAGACAGATCCTCTTAACACAAAGTCATCTGCACAA 77332
51 ValValProTyrAlaPro.ValLysArgGluGlnProLysProAlaValL 67
: : : : : : : : : : : : : : : : : : : : : : : :
77331 ATCCTCCCATGTATACCTAGTCCCAAGAAAGTCCAAGGCTTTAGCAA 77282
67 ysGlnAspGluGlnLysProLysArgGlnAlaSerHisTrpAlaValLys 83
: : : : : : : : : : : : : : : : : : : : : : : :
77281 GACATTC.....AGG 77271
84 ProThrAlaValGlyValHisValProLeuProLysLysGlnGluAlaLe 100
: : : : : : : : : : : : : : : : : : : : : : : :
77270 CCCTTGGCAGCGGATCCTTCCCTCGTCCGCGCTCATCGCTCAGCCT 77221
100 uGluPro 102
: : : : :
77220 CCCTCTCT 77214

seq_name: gb_htg14:AC044788
seq_documentation_block:
LOCUS AC044788 128305 bp DNA HTG 19-AUG-2000
DEFINITION Homo sapiens chromosome 15 clone CTD-2329A4 map 15q15, WORKING
DRAFT SEQUENCE, 7 ordered pieces.
ACCESSION AC044788
VERSION AC044788.4 GI:9857502
KEYWORDS HTG: HTGS_PHASE2; HTGS_DRAFT.
SOURCE human.

```

```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 128305)
AUTHORS Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G., Kaur,A., Madan,A.,
Nesbitt,R., Traicoff,R. and Hood,L.
TITLE Sequencing of human chromosome 15 D15S146-D15S117 region
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 128305)
AUTHORS Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B.,
Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G.,
James,R., Kaur,A., Madan,A., Owen,M.P., Ratcliffe,A., Shaffer,T.
and Hood,L.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) Multimegabase Sequencing Center, University
of Washington, PO BOX 357730, Seattle, WA 98195, USA
COMMENT On Aug 19, 2000 this sequence version replaced gi:8347945.
----- Genome Center
Center: Multimegabase Sequencing Center
Center code: UWMSC
Web site: http://chroma.mbt.washington.edu/msg_www
Contact: leerowen@systemsbiology.org
----- Summary Statistics
Sequencing vector: pUC18; L08752
Chemistry: Dye-terminator Big Dye; 90% of reads
Chemistry: Dye-terminator Big Dye; 10% of reads
Assembly program: Phrap; version 0.990399
Insert size: 120000; agarose-fp
Quality coverage: 7.5x in Q20 bases; sum-of-contigs
-----
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
1 49805: contig of 49805 bp in length
* 49806 49905: gap of unknown length
* 49906 63175: contig of 13270 bp in length
* 63176 63275: gap of unknown length
* 63276 72578: contig of 9303 bp in length
* 72579 72678: gap of unknown length
* 72679 104377: contig of 31699 bp in length
* 104378 105439: contig of 962 bp in length
* 105440 105539: gap of unknown length
* 105540 110924: contig of 5385 bp in length
* 110925 111024: gap of unknown length
* 111025 128305: contig of 17281 bp in length.
*
Location/Qualifiers
1. 128305
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15q15"
/clone="CTD-2329A4"
/clone_lib="Cal Tech human BAC library D"
/note="This clone overlaps RP11-265N6"
BASE COUNT 33339 a 27068 c 28107 g 38637 t 1154 others

```


This Page Blank (uspto)

OM of: US-09-301-906-5 to: N_Geneseq_36.* out_format : pfs

Date: Feb 14, 2001 3:10 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame-p2n.model -DEV=xlp
-O=/cgn2_1/USPTO_spool/US09301906/runat_31012001_152936_16849/app_query.fasta_1.271
-DB=N_Geneseq_36 -QFMT=fastap -SUFFIX=ring -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEX=7.000 -START=1 -MATRIX=blomsum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09301906_CGNI_1_121 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPY
-WAIT -THREADS=1

Search information block:

Query: US-09-301-906-5

Query length: 120

Database: N_Geneseq_36.*

Database sequences: 480022

Database length: 18781343

Search time (sec): 98.530000

score_list:

Sequence	Strd Orig	Zscore	EScore	len	Documentation
/SID56/gcgdata/geneseq/geneseq/NA2000.DAT:Z49203 +			605.00	1313.13	2.4e-65
/SID56/gcgdata/geneseq/geneseq/NA2000.DAT:Z49203 +			605.00	1313.13	360
/SID56/gcgdata/geneseq/geneseq/NA2000.DAT:Z49208 +			605.00	1281.54	6714
/SID56/gcgdata/geneseq/geneseq/NA2000.DAT:Z49200 +			605.00	1270.94	17919
/SID56/gcgdata/geneseq/geneseq/NA1998.DAT:V52204 +			88.00	132.84	32768
/SID56/gcgdata/geneseq/geneseq/NA1998.DAT:V52496 +			86.00	149.95	1.46
/SID56/gcgdata/geneseq/geneseq/NA1998.DAT:V52503 +			85.00	167.07	0.1619
/SID56/gcgdata/geneseq/geneseq/NA1998.DAT:V52501 +			85.00	164.62	0.2217
/SID56/gcgdata/geneseq/geneseq/NA1998.DAT:V52506 +			84.00	161.06	0.3498
/SID56/gcgdata/geneseq/geneseq/NA1999.DAT:X13101 +			83.00	123.85	41.36
/SID56/gcgdata/geneseq/geneseq/NA2000.DAT:Z29679 -			80.50	136.43	8.24
/SID56/gcgdata/geneseq/geneseq/NA1994.DAT:Q63499 +			80.50	129.87	19.10
/SID56/gcgdata/geneseq/geneseq/NA1998.DAT:V52508 +			80.00	147.27	2.05
/SID56/gcgdata/geneseq/geneseq/NA1998.DAT:V52518 +			78.50	155.50	0.7134
/SID56/gcgdata/geneseq/geneseq/NA1999.DAT:X13065 -			76.50	107.67	329.25
/SID56/gcgdata/geneseq/geneseq/NA1999.DAT:Z32020 -			75.00	102.58	632.41
/SID56/gcgdata/geneseq/geneseq/NA2000.DAT:A05548 +			74.00	125.95	31.58
/SID56/gcgdata/geneseq/geneseq/NA1997.DAT:T85728 +			74.00	120.12	66.74
/SID56/gcgdata/geneseq/geneseq/NA1999.DAT:Z20083 +			73.50	122.41	49.72
/SID56/gcgdata/geneseq/geneseq/NA1997.DAT:T77031 -			72.50	136.23	8.45
/SID56/gcgdata/geneseq/geneseq/NA1992.DAT:Q27189 +			72.50	119.95	68.18
/SID56/gcgdata/geneseq/geneseq/NA1996.DAT:T27081 +			72.00	136.55	8.11
/SID56/gcgdata/geneseq/geneseq/NA1999.DAT:Z12253 +			71.50	118.40	83.24
/SID56/gcgdata/geneseq/geneseq/NA1992.DAT:Q22871 +			71.00	110.20	238.23
/SID56/gcgdata/geneseq/geneseq/NA1998.DAT:X13978 -			71.00	139.41	5.62
/SID56/gcgdata/geneseq/geneseq/NA1998.DAT:X14033 +			71.00	133.78	11.57
/SID56/gcgdata/geneseq/geneseq/NA1997.DAT:T68824 -			71.00	130.68	17.23
/SID56/gcgdata/geneseq/geneseq/NA1999.DAT:X9557 +			71.00	123.97	40.71
/SID56/gcgdata/geneseq/geneseq/NA1998.DAT:V20067 +			71.00	123.27	44.53
/SID56/gcgdata/geneseq/geneseq/NA1998.DAT:V52497 +			71.00	123.21	44.90
/SID56/gcgdata/geneseq/geneseq/NA1996.DAT:T30869 +			70.50	123.69	42.19
/SID56/gcgdata/geneseq/geneseq/NA2000.DAT:Z38917 +			70.50	116.31	108.77
/SID56/gcgdata/geneseq/geneseq/NA1999.DAT:X13275 +			70.50	104.40	501.19
/SID56/gcgdata/geneseq/geneseq/NA1999.DAT:V87444 -			70.00	138.59	6.24
/SID56/gcgdata/geneseq/geneseq/NA1995.DAT:Q76024 +			70.00	118.23	85.07
/SID56/gcgdata/geneseq/geneseq/NA1991.DAT:Q10529 +			70.00	110.74	222.12
/SID56/gcgdata/geneseq/geneseq/NA1992.DAT:Q25204 +			69.50	128.21	23.64
/SID56/gcgdata/geneseq/geneseq/NA1997.DAT:T86166 +			69.50	128.21	23.64
/SID56/gcgdata/geneseq/geneseq/NA1992.DAT:Q25815 +			69.50	115.06	127.61
/SID56/gcgdata/geneseq/geneseq/NA1994.DAT:Q74770 +			69.50	112.16	185.29
/SID56/gcgdata/geneseq/geneseq/NA1994.DAT:Q74770 +			69.50	105.72	422.96

/SID56/gcgdata/geneseq/geneseq/NA2000.DAT:Z97038 - 69.00 121.57 55.41
/SID56/gcgdata/geneseq/geneseq/NA1999.DAT:Z31957 + 69.00 119.72 70.25
/SID56/gcgdata/geneseq/geneseq/NA1996.DAT:T39799 + 69.00 117.54 92.85
/SID56/gcgdata/geneseq/geneseq/NA1997.DAT:T84563 + 69.00 117.27 96.13
/SID56/gcgdata/geneseq/geneseq/NA1997.DAT:T43379 + 69.00 117.27 96.13

seq_name: /SID56/gcgdata/geneseq/geneseq/NA2000.DAT:Z49203

seq_documentation_block:

ID Z49203 standard; DNA; 360 BP.

AC Z49203;

XX 07-MAR-2000 (first entry)

DT GLRAV-3 polyprotein proteinase domain DNA.

DE Grapevine; leafroll; grapevine leafroll-associated virus 3; GLRAV-3;
KW viral disease; yield loss; sugar content; inhibition; infection;
KW replication; polyprotein; domain; proteinase; methyltransferase;
KW helicase; RNA-dependent; RNA polymerase; untranslated region;
KW transgenic plant; component; resistant; truncation; deletion; antisense;
KW expression; detection; antibody; ds.

OS Grapevine leafroll-associated virus 3.

XX Key Location/Qualifiers

FH mat_peptide 1..360

FT /*tag= a

FT /product= "GLRAV-3 polyprotein proteinase domain".

XX WO9955880-A1.

XX 04-NOV-1999.

XX 29-APR-1999; 99WO-US09307.

XX 29-APR-1998; 98US-0083404.

XX (CORR) CORNELL RES FOUND INC.

XX Gonsalves D, Ling K;

XX WPI; 2000-062035/05.

XX P-PSDB; Y58143.

XX Newly Isolated grapevine leafroll virus protein or polypeptide useful
for producing transgenic plants conferring viral disease resistance -

XX Claim 12; Fig 5; 84pp; English.

This sequence represents DNA encoding the proteinase domain of the
polyprotein (Y58148) from the grapevine leafroll-associated virus 3
(GLRAV-3). Leafroll is a serious viral disease, occurring wherever
grapes are grown. Although not lethal, it causes yield losses and
reduction in sugar content. The virus encodes several proteins,
which may serve as targets for the inhibition of viral infection or
replication. These proteins include the 242-248 kD polyprotein (Y58148),
encoded by open reading frame (ORF) 1a, and which comprises a proteinase
domain (Y58143), a methyltransferase domain (Y58144) and a helicase
domain (Y58145); a RNA-dependent RNA polymerase (Y58146), encoded by ORF
1b; and a protein of unspecified function (Y58147), encoded by ORF 11.
Nucleotides encoding these proteins, or fragments thereof, and the 5',
and 3' untranslated regions (UTRs) of the genome (249201-249202) are
useful for the generation of transgenic plants and plant components.
Such transgenic plants may be resistant to viral disease, for example,
this property being conferred on the plants via the use of nucleotides
encoding truncated or internally deleted proteins, or via the use of
antisense nucleotides to inhibit viral gene expression. The nucleotides
may additionally be used in the detection of viral nucleic acids in a
tissue sample. The proteins can be used to generate antibodies which
may be used to detect GLRAV-3 in plant samples. The isolation of
GLRAV-3 facilitates the production of agents that reduce the risk of

```
CC infection or damage by the virus in vineyards.
XX Sequence 360 BP; 90 A; 97 C; 100 G; 73 T; 0 other;
SQ

alignment_scores:
  Quality: 605.00      Length: 120
  Ratio: 5.042        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-301-906-5 x Z49203 ..

Align seg 1/1 to: Z49203 from: 1 to: 360

1 ValSerGlySerValSerAlaLeuArgGlyAspGlyLysLysValLeuMe 17
1 GTcAGCGGCTcAGTcAGCGGCTgAGAGGGATgGTAAGAAGGTCTTGAT 50
17 tGluAlaArgThrSerThrSerAlaThrSerAspValSerAspPheAspV 34
51 GGAGGAAGGACCTCAACTCCGCACTTCGACGCTGCTGTGTTTCGACG 100
34 alValPheGluAlaValSerAsnAlaLeuValValHisTyrHisArg 50
101 TCGTATTTCGAAGCTGTTCTTAATGCAATTACTTGTCTACACTACCCCG 150
51 ValValProTyrAlaProValLysArgGluInProLysProAlaValLy 67
151 GTAGTCGGTATGCCCGGTCAGCGGAGCAGCCCTAAACCGCGCTGTAA 200
67 sGInAspGluGlnLysProLysArgGlnAlaSerHisTrpAlaValLysP 84
201 GCAAGATGACGACAGACCCCAACGCGAGCGTCACATTGGCGCTGTTAAGC 250
84 roThrAlaValGlyValHisValProLeuProLysLysGlnGluAlaLeu 100
251 CAACAGCTGTTCGGTCCAGCTACCACTTCTCTATAAACAAGGAGCACTG 300
101 GluProAlaGlnSerValProGlnGlnSerLeuGluGluLysAlaAlaLe 117
301 GAGCGAGCGGCAATCAGTCCCAACAACAGTCTGTGGAGGAGGAGCGCCCT 350
117 uThrPheGly 120
351 GACGTTTGGC 360

seq_name: /SIDS6/gcgdata/geneseq/geneseq/NA2000.DAT:Z49208
seq_documentation_block:
ID Z49208 standard; DNA: 6714 BP.
XX
AC Z49208;
XX
DT 07-MAR-2000 (first entry)
XX
DE GLRaV-3 ORF 1a, encoding polyprotein.
XX
KW Grapevine: leafroll; grapevine leafroll-associated virus 3; GLRaV-3;
KW viral disease; yield loss; sugar content; inhibition; infection;
KW replication; polyprotein; domain; proteinase; methyltransferase;
KW helicase; RNA-dependent; RNA polymerase; untranslated region;
KW transgenic plant; component; resistant; truncation; deletion; antisense;
KW expression; detection; antibody; ds.
XX
OS Grapevine leafroll-associated virus 3.
XX
FH Key Location/Qualifiers
FT CDS 1..6714
FT FT /*tag= a
FT FT /note= "ORF 1a"
FT FT /product= "GLRaV-3 polyprotein"
FT FT mat_peptide 253..612
```

```
FT /*tag= b
FT /product= "GLRaV-3 polyprotein proteinase domain
FT (Y58143)"
FT mat_peptide 1378..2193
FT /*tag= c
FT /product= "GLRaV-3 polyprotein methyltransferase domain
FT (Y58144)"
FT mat_peptide 5764..6636
FT /*tag= d
FT /product= "GLRaV-3 polyprotein helicase domain (Y58145)"
XX
PN WO9955880-A1.
XX
XX 04-NOV-1999.
XX
XX 29-APR-1999; 99WO-US09307.
XX
XX 29-APR-1998; 98US-0083404.
XX
XX (CORR ) CORNELL RES FOUND INC.
XX
XX Gonsalves D, Ling K;
XX
XX WPI; 2000-062035/05.
DR P-PSDB; Y58148.
XX
PT Newly isolated grapevine leafroll virus protein or polypeptide useful
PT for producing transgenic plants conferring viral disease resistance -
XX
XX Claim 7; Fig 4; 84pp; English.
XX
CC This sequence represents open reading frame (ORF) 1a which encodes
CC the polyprotein of the grapevine leafroll-associated virus 3
CC (GLRaV-3). Leafroll is a serious viral disease, occurring wherever
CC grapes are grown. Although not lethal, it causes yield losses and
CC reduction in sugar content. The virus encodes several proteins,
CC which may serve as targets for the inhibition of viral infection or
CC replication. These proteins include the 242-248 kD polyprotein (Y58148),
CC encoded by open reading frame (ORF) 1a, and which comprises a proteinase
CC domain (Y58143), a methyltransferase domain (Y58144) and a helicase
CC domain (Y58145); a RNA-dependent RNA polymerase (Y58146), encoded by ORF
CC 1b; and a protein of unspecified function (Y58147), encoded by ORF 11.
CC Nucleotides encoding these proteins, or fragments thereof, and the 5',
CC and 3' untranslated regions (UTRs) of the genome (Z49201-Z49202) are
CC useful for the generation of transgenic plants and plant components.
CC Such transgenic plants may be resistant to viral disease, for example,
CC this property being conferred on the plants via the use of nucleotides
CC encoding truncated or internally deleted proteins, or via the use of
CC antisense nucleotides to inhibit viral gene expression. The nucleotides
CC may additionally be used in the detection of viral nucleic acids in a
CC tissue sample. The proteins can be used to generate antibodies which
CC may be used to detect GLRaV-3 in plant samples. The isolation of
CC GLRaV-3 facilitates the production of agents that reduce the risk of
CC infection or damage by the virus in vineyards.
XX
XX Sequence 6714 BP; 1675 A; 1429 C; 1877 G; 1733 T; 0 other;
SQ

alignment_scores:
  Quality: 605.00      Length: 120
  Ratio: 5.042        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-301-906-5 x Z49208 ..

Align seg 1/1 to: Z49208 from: 1 to: 6714

1 ValSerGlySerValSerAlaLeuArgGlyAspGlyLysLysValLeuMe 17
253 GTcAGCGGCTcAGTcAGCGGCTgAGAGGGATgGTAAGAAGGTCTTGAT 302
17 tGluAlaArgThrSerThrSerAlaThrSerAspValSerAspPheAspV 34
```

```

|||||
303 CGAGCAAGGACCTCAACTCCGCAACTTCCGACGTCTGATTTCACG 352
34 alValPheGluAlaValSerAsnAlaLeuLeuValValHisTyrHisArg 50
|||||
353 TCGTATTGGAAGCTGTTCTAATGCATTACTTGTGCTACACTACCACGG 402
51 ValValProTyrAlaProValLysArgGluGlnProLysProAlaValLy 67
403 GTAGTGGCGTATGCCCGCTCAAGCGGAGCAGCGCTAAACGGCTGTAA 452
67 sClnAspGluClnLysProLysArgGlnAlaSerHisTrpAlaValLysp 84
|||||
453 GCAAGATGAGCAGAGCCCAACCGCAGCGTCACATTGGCTGTTAAGC 502
84 roThrAlaValGlyValHisValProLeuProLysLysGlnGluAlaLeu 100
|||||
503 CAACAGCTGTTGGCGTCACGTACCACTTCCTAAAAAAGAGGAGCACTG 552
101 GluProAlaGlnSerValProGlnGlnSerLeuGluGluLysAlaAlaLe 117
553 GAGCCAGCGCAATCAGTCCCAACACAGTCGTTGGAGGAGGAGCGGCTT 602
117 uThrPheGly 120
|||||
603 GACGTTTGGC 612

```

seq_name: /SIDS6/gcgdata/geneseq/geneseqn/NA2000.DAT:Z49200

seq_documentation_block:

XX 249200 standard; DNA; 17919 BP.

XX AC 249200;

XX DT 07-MAR-2000 (first entry)

XX Grapevine leafroll-associated virus 3 (GLRaV-3) genome.

XX Grapevine; leafroll; grapevine leafroll-associated virus 3; GLRaV-3;
 KW viral disease; yield loss; sugar content; inhibition; infection;
 KW replication; polyprotein; domain; proteinase; methyltransferase;
 KW helicase; RNA-dependent; RNA polymerase; untranslated region;
 KW transgenic plant; component; resistant; truncation; deletion; antisense;
 KW expression; detection; antibody; ds.

XX Grapevine leafroll-associated virus 3.

```

XX Key Location/Qualifiers
XX 5'UTR 1..158
XX /tag= a
XX CDS 159..6872
XX /tag= b
XX /note= "ORF 1a"
XX /product= "GLRaV-3 polyprotein (Y58148)"
XX 411..770
XX /tag= c
XX /product= "GLRaV-3 polyprotein proteinase domain
  (Y58143)"
XX 1536..2351
XX /tag= d
XX /product= "GLRaV-3 polyprotein methyltransferase domain
  (Y58144)"
XX 5922..6794
XX /tag= e
XX /product= "GLRaV-3 polyprotein helicase domain (Y58145)"
XX 6877..8478
XX /tag= f
XX /note= "ORF 1b"
XX /product= "GLRaV-3 RNA-dependent RNA polymerase (Y58146)"
XX 8708..8863
XX /tag= g
XX /note= "ORF 2"
XX 9930..10067
XX CDS

```

```

FT /tag= h
FT /note= "ORF 3"
FT 10086..11735
FT /tag= i
FT /note= "ORF 4"
FT 11728..13179
FT /tag= j
FT /note= "ORF 5"
FT 13269..14210
FT /tag= k
FT /note= "ORF 6"
FT 14273..15706
FT /tag= l
FT /note= "ORF 7"
FT 15717..16274
FT /tag= m
FT /note= "ORF 8"
FT 16271..16804
FT /tag= n
FT /note= "ORF 9"
FT 16811..17350
FT /tag= o
FT /note= "ORF 10"
FT 17353..17463
FT /tag= p
FT /note= "ORF 11"
FT /product= "GLRaV-3 ORF 11-encoded protein (Y58147)"
FT 17460..17642
FT /tag= q
FT /note= "ORF 12"
FT 17643..17919
FT /tag= r
FT 3'UTR
FT
FT WO9955880-A1.
FT
FT 04-NOV-1999.
FT
FT 29-APR-1999; 99WO-US09307.
FT
FT 29-APR-1998; 98US-0083404.
FT (CORR ) CORNELL RES FOUND INC.
FT
FT Gonsalves D, Ling K;
FT
FT WPI; 2000-062035/05.
FT P-PSDB; Y58143, Y58144, Y58145, Y58146, Y58147, Y58148.
FT
FT Newly isolated grapevine leafroll virus protein or polypeptide useful
  for producing transgenic plants conferring viral disease resistance -
  Example 1; Fig 2; 84pp; English.

```

This sequence represents the genome of grapevine leafroll-associated virus 3 (GLRaV-3). Leafroll is a serious viral disease, occurring wherever grapes are grown. Although not lethal, it causes yield losses and reduction in sugar content. The virus encodes several proteins, which may serve as targets for the inhibition of viral infection or replication. These proteins include a 242-248 kD polyprotein (Y58148), encoded by open reading frame (ORF) 1a, and which comprises a proteinase domain (Y58143), a methyltransferase domain (Y58144) and a helicase domain (Y58145); a RNA-dependent RNA polymerase (Y58146), encoded by ORF 1b; and a protein of unspecified function (Y58147), encoded by ORF 11. Nucleotides encoding these proteins, or fragments thereof, and the 5' and 3' untranslated regions (UTRs) of the genome (Z49201-Z49202) are useful for the generation of transgenic plants and plant components. Such transgenic plants may be resistant to viral disease, for example, this property being conferred on the plants via the use of nucleotides encoding truncated or internally deleted proteins, or via the use of antisense nucleotides to inhibit viral gene expression. The nucleotides may additionally be used in the detection of viral nucleic acids in a tissue sample. The proteins can be used to generate antibodies which may be used to detect GLRaV-3 in plant samples. The isolation of

CC GLRAV-3 facilitates the production of agents that reduce the risk of
CC infection or damage by the virus in vineyards.

XX
SQ Sequence 17919 BP; 4729 A; 3527 C; 4740 G; 4923 T; 0 other;

alignment_scores:
Quality: 605.00 Length: 120
Ratio: 5.042 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-301-906-5 x 249200 ..

Align seg 1/1 to: 249200 from: 1 to: 17919

1 ValSerGlySerValSerAlaLeuArgGlyAspGlyLysValLeuMe 17
|||||
411 GTCAGCGGCTCAGTCAGCGGCTGAGAGGGATGGTAAGAAGCTCTTGAT 460
17 tGluAlaArgThrSerThrSerAlaThrSerAspValSerAspPheAspV 34
|||||
461 GGAGGCAGAGACTCAACTTCCGCACCTCCGACGTCGTGATTTGACG 510
34 aValPheGluAlaValSerAsnAlaLeuValValHisTyrHisArg 50
|||||
511 TCGTATTGGAAGCTGTTCTTAATGCAATTACTTTCGTACACTACCCGG 560
51 ValValProTyrAlaProValLysArgGluGlnProLysProAlaVally 67
|||||
561 GTATGTCGCTATCCCGCTCAAGCGCGAGCAGCCCTAAACCGCTGTAA 610
67 sGlnAspGluGlnLysProLysArgGlnAlaSerHisTrpAlaValLysP 84
|||||
611 GCAGATGACAGAGAGCCCAAGCGGAGCGTCACATTTGGCTGTAAAG 660
84 roThrAlaValGlyValHisValProLeuProLysLysGlnGluAlaLeu 100
|||||
661 CAACAGCTGTTGGCTCCAGCTACCACTTCTTAAACAGAGGAGCACTG 710
101 GluProAlaGlnSerValProGlnGlnSerLeuGluGluLysAlaLale 117
|||||
711 GAGCCAGCGCAATTCAGTCCACACACAGCTGTTGGAGGAGAGCCGCTT 760
117 uThrPheGly 120
|||||
761 GAGCTTTGGC 770

seq_name: /SID56/gcgdata/geneseq/geneseqn/NA1998.DAT:V52204

seq_documentation_block:

ID V52204 standard; DNA; 32768 BP.

XX V52204;

AC V52204;

XX 23-OCT-1998 (first entry)

DE Streptococcus pneumoniae genome fragment SEQ ID NO:71.

XX Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW computer readable medium; vaccine; pharmaceutical composition; ds.

XX Streptococcus pneumoniae.

OS WO9818931-A2.

PN 07-MAY-1998.

XX 30-OCT-1997; 97WO-US19588.

XX 31-OCT-1996; 96US-0029960.

XX (HUMA-) HUMAN GENOME SCI INC.

XX

PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;

XX Kunsch CA, Rosen CA;

XX WPI; 1998-272225/24.

DR

XX Computer-readable medium with recorded Streptococcus pneumoniae
PT polynucleotide sequences - useful in diagnostic kits and assays, and
PT pharmaceutical compositions and vaccines for Streptococcus
PT pneumoniae

XX Claim 1; Page 575-593; 1409pp; English.

XX The present invention describes a computer readable medium which has
CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded
CC on it, or a representative fragment or a sequence at least 95% identical
CC to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus
CC pneumoniae. The present invention also describes an isolated nucleic acid
CC molecule encoding a homologue of any of the fragments of the S. pneumoniae
CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
CC by a process comprising: (a) screening a genomic DNA library using as a
CC probe a target sequence defined by any of the sequences in SEQ ID NO:1
CC to 391, identifying members of the library which contain sequences
CC that hybridize to the target sequence and isolating the nucleic acid
CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced
CC from an organism amplifying nucleic acid molecules whose nucleotide
CC sequence is homologous to amplification primers derived from the
CC fragment of the S. pneumoniae genome to prime the amplification and
CC isolating the amplified sequences. The computer readable medium can be
CC used in a computer-based system for identifying fragments of the
CC S. pneumoniae genome of commercial importance, or expression modulating
CC fragments of the S. pneumoniae genome. Products from the present
CC invention can be used in diagnosis kits and assays, and pharmaceutical
CC compositions and vaccines for S. pneumoniae.

XX Sequence 32768 BP; 9834 A; 6168 C; 7154 G; 9609 T; 3 other;

alignment_scores:

Quality: 88.00 Length: 150
Ratio: 1.158 Gaps: 6
Percent Similarity: 50.667 Percent Identity: 27.333

alignment_block:

US-09-301-906-5 x V52204 ..

Align seg 1/1 to: V52204 from: 1 to: 32768

2 SerGlySerValSerAlaLeuArgGlyAspGlyLysValLeuMetG1 18
|||||
5363 TCAGCGTCAACAGTGCATCAGTCTCAGCAAGCAGCAGTCCGTCGGCTTC 5412
18 uAlaArgThrSerThrSerAlaThrSerAspValSerAspPheAspValV 35
|||||
5413 AGCATCAACGAGTGCCTCAGCCTCAGCAAGTATCTCAGCGTCTGAATCGG 5462
35 alPheGluAlaValSerAsnAlaLeu.LeuValValHisTyrHisArgVa 51
:::|
5463 CATCAACGAGTGCCTCAGCCTCAGCAAGTACTAGTCATCGG..... 5504
51 lValProTyrAlaProValLysArgGluGlnProLysProAlaValLysG 68
:::|
5505 .CTTCAGCANGCAGTGCCTGCGGCTTCAGCATCAACCA....GTGCCTC 5550
68 lnAspGluGlnLysProLysArgGlnAlaSerHisTrpAlaValLys... 83
|||
5551 AGCCTCAGCAAGTATCTCAGCGTCTGAATCGCATCAACGAGTCCGCTCAG 5600
84 ProThrAlaValGlyValHis.....
|||
5601 CCTCAGCAAGTACTAGTGCATCAGCMTCACATCAACGAGTGCATCGGCT 5650


```

PN WO9818931-A2.
XX
PD
PP
XX 07-MAY-1998.
XX
XX 30-OCT-1997; 97WO-US19588.
XX
XX 31-OCT-1996; 96US-0029960.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;
XX Kunsch CA, Rosen CA;
XX WPI; 1998-272225/24.
XX
XX Computer-readable medium with recorded Streptococcus pneumoniae
XX polynucleotide sequences - useful in diagnostic kits and assays, and
XX pharmaceutical compositions and vaccines for Streptococcus
XX pneumoniae
XX
XX Claim 1; Page 1384-1385; 1409pp; English.
XX
XX The present invention describes a computer readable medium which has
XX the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded
XX on it, or a representative fragment or a sequence at least 95% identical
XX to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
XX to 391 (V52134 to V52524) are genomic fragments from Streptococcus
XX pneumoniae. The present invention also describes an isolated nucleic acid
XX molecule encoding a homologue of any of the fragments of the S.pneumoniae
XX genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
XX by a process comprising: (a) screening a genomic DNA library using as a
XX probe a target sequence defined by any of the sequences in SEQ ID NO:1
XX to 391, identifying members of the library which contain sequences
XX that hybridise to the target sequence and isolating the nucleic acid
XX molecules from the members; or (b) isolating mRNA, DNA or cDNA produced
XX from an organism, amplifying nucleic acid molecules whose nucleotide
XX sequence is homologous to amplification primers derived from the
XX fragment of the S. pneumoniae genome to prime the amplification and
XX isolating the amplified sequences. The computer readable medium can be
XX used in a computer-based system for identifying fragments of the
XX S. pneumoniae genome of commercial importance, or expression modulating
XX fragments of the S. pneumoniae genome. Products from the present
XX invention can be used in diagnosis kits and assays, and pharmaceutical
XX compositions and vaccines for S. pneumoniae.
XX
XX Sequence 750 BP; 195 A; 238 C; 173 G; 144 T; 0 other;
XX
alignment_scores:
  Quality: 85.00 Length: 134
  Ratio: 1.118 Gaps: 7
  Percent Similarity: 56.716 Percent Identity: 30.597
alignment_block:
  US-09-301-906-5 x V52503 ..
  Align seg 1/1 to: V52503 from: 1 to: 750
  2 SerClySerValSerAlaLeuArgGlyAspGlyLysLysValLeuMetG1 18
  |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
  163 TCAGCGTCACCAAGTGTGATCAGTCTCAGCAAGCACCAGCGTGGGCTC 212
  ||| |||||:||||:||||:||||:||||:||||:||||:||||:
  18 uAlaArgThrSerThrSerAlaThrSerAspValSerAspPheAspValV 35
  :||| |||||:||||:||||:||||:||||:||||:||||:||||:
  213 AGCATCAACAGATGCCCTCAGCCCTCAGCAAGTATCTCAGCGTCTGATCGG 262
  :|||:||||:||||:||||:||||:||||:||||:||||:||||:
  35 alPheGluAlaValSerAsnAlaLeu.LeuValValHisTyrHisArgVa 51
  :|||:||||:||||:||||:||||:||||:||||:||||:||||:
  263 CATCAACGAGTGGCTCAGCTCAGCAAGTACTAGTGCATCGG..... 304
  ||| |||||:||||:||||:||||:||||:||||:||||:||||:
  51 lValProTyrAlaProValLysArgGluInProLysProAlaValLysG 68
  :||| |||||:||||:||||:||||:||||:||||:||||:||||:
  305 .CTTCAGCAAGCACCAGTGGCTGGCTTTCAGCATCAACCA...GTGCCTC 350

```

```

68 lnAspGluGlnLysProLysArgGlnAlaSerHisTrpAlaValLys... 83
|||:||||:||||:||||:||||:||||:||||:||||:||||:
351 AGCCTCAGCAAGTATCTCAGCGTCTGAATCGCATCAACGAGTGGGTCAG 400
||| |||||:||||:||||:||||:||||:||||:||||:
84 ProThrAlaValGlyValHis..... 90
||| |||||:||||:||||:||||:||||:||||:||||:
401 CCTCAGCAAGTACTAGTGCATCAGCATCAGCATCAACGAGTGCATCGGCT 450
||| |||||:||||:||||:||||:||||:||||:||||:
91 .....ValProLeuProLysLysGlnGluAla...LeuGluProAlaG 104
||||| |||||:||||:||||:||||:||||:||||:||||:
451 TCAGCAAGTACCAGCGCCTCAGCTTCAGCAAGCACCAGTGCCTCAGCTTC 500
||| |||||:||||:||||:||||:||||:||||:||||:
104 lnSerValPro.GlnGlnSerLeuGluGlnLysAlaAlaLeuThrPhe 119
|||:|||||:||||:||||:||||:||||:||||:||||:
501 AGCAAGTACCAGCGCCTCAGCTCAGCAAGCACCAGTGCCTCAGCTTC 548
||| |||||:||||:||||:||||:||||:||||:||||:
seq_name: /SID6/gcgdata/geneseq/geneseqn/NA1998.DAT:V52501

```

```

seq_documentation_block:
ID V52501 standard; DNA; 941 BP.
XX
AC V52501;
XX
DT 23-OCT-1998 (first entry)
XX
DE Streptococcus pneumoniae genome fragment SEQ ID NO:368.
XX
KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
XX computer readable medium; vaccine; pharmaceutical composition; ds.
XX
OS Streptococcus pneumoniae.
XX
PN WO9818931-A2.
XX
PD 07-MAY-1998.
XX
PF 30-OCT-1997; 97WO-US19588.
XX
PR 31-OCT-1996; 96US-0029960.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;
XX Kunsch CA, Rosen CA;
XX WPI; 1998-272225/24.
XX
PT Computer-readable medium with recorded Streptococcus pneumoniae
XX polynucleotide sequences - useful in diagnostic kits and assays, and
XX pharmaceutical compositions and vaccines for Streptococcus
XX pneumoniae
XX
PS Claim 1; Page 1383-1384; 1409pp; English.
XX
XX The present invention describes a computer readable medium which has
XX the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded
XX on it, or a representative fragment or a sequence at least 95% identical
XX to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
XX to 391 (V52134 to V52524) are genomic fragments from Streptococcus
XX pneumoniae. The present invention also describes an isolated nucleic acid
XX molecule encoding a homologue of any of the fragments of the S.pneumoniae
XX genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
XX by a process comprising: (a) screening a genomic DNA library using as a
XX probe a target sequence defined by any of the sequences in SEQ ID NO:1
XX to 391, identifying members of the library which contain sequences
XX that hybridise to the target sequence and isolating the nucleic acid
XX molecules from the members; or (b) isolating mRNA, DNA or cDNA produced
XX from an organism, amplifying nucleic acid molecules whose nucleotide
XX sequence is homologous to amplification primers derived from the
XX fragment of the S. pneumoniae genome to prime the amplification and
XX isolating the amplified sequences. The computer readable medium can be
XX used in a computer-based system for identifying fragments of the

```


alignment_scores:
 Quality: 80.50 Length: 90
 Ratio: 1.830 Gaps: 3
 Percent Similarity: 48.889 Percent Identity: 26.667

alignment_block:
 US-09-301-906-5 x 229679/rev ..

Align seg 1/1 to reverse of: 229679 from: 1 to: 5140

```

51 ValValProTyraAlaProValLysArgGlu..... 60
||||| ||||| |||||
2982 GTACTGAGCTTAGCCCATCGCAGCGCACCTTCGGCGCCTACTTGCAC 2933

61 .GlnProLysProAlaValLysGlnAspGluGlnLysProLysArgGlnA 77
||||| ||||| ||||| ||||| |||||
2932 GAGGCTCGCGCTTCGGTTGCCAAAGCCATCAGCCGCCAGCGCAATG 2883

77 laSerHisTrpAlaValLysProThrAlaValGlyValHis..... 90
||||| ||||| ||||| ||||| |||||
2882 CCAGACACAACTCAACAACCGATGCAATCGGAACGTACTGAGTTAG 2833

91 ..... 91
||||| ||||| ||||| ||||| |||||
2832 CCCAATGCGCAGCGCACTTCGGCGCTACTTGCACGAGCGCTTCGCGT 2783

91 lProLeuProLysLysGlnGluAlaLeuGlu.....ProAlaGlns 105
||||| ||||| ||||| ||||| |||||
2782 TCCGTTGCCAAAGCCTATCAGCGCCAGCAGCGATGTCGCCCTCCGAT 2733

105 erValProGlnGlnSerLeu 111
||||| ||||| ||||| ||||| |||||
2732 GTTCCCGCCACAAATCGATG 2713

```

seq_name: /SID56/gcgdata/geneseq/geneseqn/NA1994.DAT:Q63499

seq_documentation_block:
 ID Q63499 standard; cDNA; 9436 BP.
 AC Q63499;
 DT 17-JAN-1995 (first entry)
 DE Blood transmissible NANBHV genome.
 XX
 KW Polymerase chain reaction; PCR; amplify; primer; non-A, non-B hepatitis;
 KW NANBH; virus; blood transmissible; detection; hepatitis virus; RT-PCR;
 KW C100 antibody; HCV RNA; NS5 region; ds.
 XX
 OS Non-A, non-B hepatitis virus.
 XX
 FH Key Location/Qualifiers
 FT CDS 342..9374
 FT /*tag= a
 XX
 PN JP06105690-A.
 XX
 PD 19-APR-1994.
 XX
 XX 10-MAR-1992; 92JP-0051885.
 XX
 PR 10-MAR-1992; 92JP-0051885.
 XX
 XX (KAEN/) KAENNO K.
 PA
 XX
 DR WPI: 1994-163130/20.
 DR P-PSDB; R53417.
 XX
 PT Blood-transmissible non-A non-B hepatitis virus DNA - used for
 PT detection of hepatitis virus
 XX
 PS Claim 1; Page 8-20; 22pp; Japanese.
 XX

CC This sequence represents the genome of a blood transmissible non-A,
 CC non-B hepatitis (NANBH) virus. This sequence was isolated using the
 CC primers given in Q63500-35. The amplified fragments are used in the
 CC detection of hepatitis virus. This target DNA was isolated from serum
 CC of chronically infected NANBH patients who were C100 antibody-positive
 CC and HCV RNA (NS5 region) positive. Reverse transcription-PCR and PCR
 CC were performed on cDNA and the total human NANBH DNA was constructed
 CC from 23 clones.
 XX

SQ Sequence 9436 BP; 1876 A; 2840 C; 2655 G; 1974 T; 91 other;

alignment_scores:
 Quality: 80.50 Length: 123
 Ratio: 1.118 Gaps: 3
 Percent Similarity: 58.537 Percent Identity: 21.138

alignment_block:
 US-09-301-906-5 x Q63499 ..

Align seg 1/1 to: Q63499 from: 1 to: 9436

```

1 ValSerGlySerValSerAlaLeuArgGlyAspGlyLysLysValLeuMe 17
||||| ||||| ||||| ||||| |||||
7050 ATGGCGGGAACATCACCCGTTGGAGTTCAGAGATAAGTGGTAATTT 7099

17 tGluAlaArgThrSerThrSerAlaThrSerAspValSerAspPheAspV 34
||||| ||||| ||||| ||||| |||||
7100 GGACTCTTCGACCCCGCTTCGAGCGGAAGAGATGAGAGGAAGTATCCG 7149

34 alValPheGluAlaValSerAsnAlaLeuValValHisTyrHisArg 50
||||| ||||| ||||| ||||| |||||
7150 TTGCAGCAGAGATCCTCGAAATCC.....AAGAAG 7181

51 ValValProTyraAlaProVal....LysArgGluGlnProLysProAlaVa 66
||||| ||||| ||||| ||||| |||||
7182 TTCCCCCGCGTGGCCATATGGGCACGCCCGGATTACACCCCTCCACT 7231

66 lLysGlnAspGluGlnLysProLys.....ArgGlnAlaSerHisTrpA 81
||||| ||||| ||||| ||||| |||||
7232 GTTAGAGTCTCGGAAAGTCCGGACTACGTCCTCCGCGGTGTCATGGGT 7281

81 laValLysProThrAlaValGlyValHisValProLeuProLysLysGln 97
||||| ||||| ||||| ||||| |||||
7282 GCCCATTCGCGCCTTACCACGCGCCTCCAATACCGCTCCACGGAAG 7331

98 GluAlaLeuGluProAlaGlnSerValProGlnGlnSerLeuGluGlu 114
||||| ||||| ||||| ||||| |||||
7332 AGGACGGTTGTTCTGACAGAGATCCACCGTGCTTCTGCTTGGCGGAGCT 7381

114 sAlaAlaLeuThrPheGly 120
||||| ||||| |||||
7382 GGCTACTAAGACTTTCGCGC 7400

```

seq_name: /SID56/gcgdata/geneseq/geneseqn/NA1998.DAT:V52500

seq_documentation_block:
 ID V52500 standard; DNA; 1702 BP.
 AC V52500;
 DT 23-OCT-1998 (first entry)
 DE Streptococcus pneumoniae genome fragment SEQ ID NO:367.
 XX
 KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
 KW computer readable medium; vaccine; pharmaceutical composition; ds.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO9818931-A2.
 XX
 PD 07-MAY-1998.

```
XX 30-OCT-1997; 97WO-US19588.
XX PF
XX PR
XX 31-OCT-1996; 96US-0029960.
XX PA
XX (HUMA-) HUMAN GENOME SCI INC.
XX PI
XX Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;
XX Kunsch CA, Rosen CA;
XX WPI; 1998-272225/24.
XX
XX Computer-readable medium with recorded Streptococcus pneumoniae
XX polynucleotide sequences - useful in diagnostic kits and assays, and
XX pharmaceutical compositions and vaccines for Streptococcus
XX pneumoniae
XX
XX Claim 1; Page 1382-1383; 1409pp; English.
XX
XX The present invention describes a computer readable medium which has
XX the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded
XX on it, or a representative fragment or a sequence at least 95% identical
XX to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
XX to 391 (V52134 to V52524) are genomic fragments from Streptococcus
XX pneumoniae. The present invention also describes an isolated nucleic acid
XX molecule encoding a homologue of any of the fragments of the S. pneumoniae
XX genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
XX by a process comprising: (a) screening a genomic DNA library using as a
XX probe a target sequence defined by any of the sequences in SEQ ID NO:1
XX to 391, identifying members of the library which contain sequences
XX that hybridise to the target sequence and isolating the nucleic acid
XX molecules from the members; or (b) isolating mRNA, DNA or cDNA produced
XX from an organism, amplifying nucleic acid molecules whose nucleotide
XX sequence is homologous to amplification primers derived from the
XX fragment of the S. pneumoniae genome to prime the amplification and
XX isolating the amplified sequences. The computer readable medium can be
XX used in a computer-based system for identifying fragments of the
XX S. pneumoniae genome of commercial importance, or expression modulating
XX fragments of the S. pneumoniae genome. Products from the present
XX invention can be used in diagnosis kits and assays, and pharmaceutical
XX compositions and vaccines for S. pneumoniae.
XX
XX Sequence 1702 BP; 422 A; 551 C; 398 G; 331 T; 0 other;
```

```
alignment_scores:
    Quality: 80.00      Length: 109
    Ratio: 1.212        Gaps: 5
    Percent similarity: 60.550      Percent Identity: 30.275
```

```
alignment_block:
```

```
US-09-301-906-5 x V52500 ..
```

```
Align seg 1/1 to: V52500 from: 1 to: 1702
```

```
2 SerGlySerValSerAlaLeuArgGlyAspGlyLysLysValLeuMetG1 18
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
275 TCGGCTTCAGCAGTACCATCGGTGCTAGCTCAGCCTCAGCAAGTACCATCGGTC 324
18 uAlaArgThrSerThrSerAlaThrSerAspValSerAspPheAspValV 35
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
325 AGCCTCAGCGTCGCAAGTGCCTCGGCTCAACGACGTGCATCTGAATCGG 374
35 alPheGluAlaValSerAsnAlaLeu. LeuValValHisTyrHisArgVa 51
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
375 CATCAACCATCGGTGCTAGCCTCAGCAAGTACTAGTCATCAG..... 416
51 lValProTyrAlaProValLysArgGluInProLysProAlaValLysG 68
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
417 .CITCAGCATCAACGAGTGCATCGGCTTCAGCATCAACCA...GTGCATC 462
68 lNAspGluInLysProLysArgGluAlaSerHisTrpAlaValLysPro 84
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
```

```
463 AGAGTCAGCAAGTACCAGTCGCTCAGCTCCGCATCAACAAGTG...CCT 509
85 ThrAlaValGlyValHisValProLeuProLysLysGlnGluAlaLeuG1 101
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
510 CGGCTTCAGCAAGTACTAGCGCTCAGCCTCAGCCTCAGCGTCAACAAGTGCCTCA 559
101 u...ProAlaGlnSerValProGln 108
: ||| ||| :|||:|
560 GCTTCGCGCTCAACGACGCGCTCGG 584
```

```
seq_name: /SIDS6/gcgdata/geneseq/geneseqn/NA1998.DAT:V52518
```

```
seq_documentation_block:
```

```
ID V52518 standard; DNA; 586 BP.
```

```
XX AC V52518;
```

```
XX DT 23-OCT-1998 (first entry)
```

```
XX DE Streptococcus pneumoniae genome fragment SEQ ID NO:385.
```

```
XX KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
```

```
XX KW computer readable medium; vaccine; pharmaceutical composition; ds.
```

```
XX OS Streptococcus pneumoniae.
```

```
XX PN WO9818931-A2.
```

```
XX PD 07-MAY-1998.
```

```
XX PF 30-OCT-1997; 97WO-US19588.
```

```
XX PR 31-OCT-1996; 96US-0029960.
```

```
XX (HUMA-) HUMAN GENOME SCI INC.
```

```
XX PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;
```

```
XX PI Kunsch CA, Rosen CA;
```

```
XX WPI; 1998-272225/24.
```

```
XX Computer-readable medium with recorded Streptococcus pneumoniae
XX polynucleotide sequences - useful in diagnostic kits and assays, and
XX pharmaceutical compositions and vaccines for Streptococcus
XX pneumoniae
```

```
XX Claim 1; Page 1394-1395; 1409pp; English.
```

```
XX The present invention describes a computer readable medium which has
XX the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded
XX on it, or a representative fragment or a sequence at least 95% identical
XX to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
XX to 391 (V52134 to V52524) are genomic fragments from Streptococcus
XX pneumoniae. The present invention also describes an isolated nucleic acid
XX molecule encoding a homologue of any of the fragments of the S. pneumoniae
XX genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
XX by a process comprising: (a) screening a genomic DNA library using as a
XX probe a target sequence defined by any of the sequences in SEQ ID NO:1
XX to 391, identifying members of the library which contain sequences
XX that hybridise to the target sequence and isolating the nucleic acid
XX molecules from the members; or (b) isolating mRNA, DNA or cDNA produced
XX from an organism, amplifying nucleic acid molecules whose nucleotide
XX sequence is homologous to amplification primers derived from the
XX fragment of the S. pneumoniae genome to prime the amplification and
XX isolating the amplified sequences. The computer readable medium can be
XX used in a computer-based system for identifying fragments of the
XX S. pneumoniae genome of commercial importance, or expression modulating
XX fragments of the S. pneumoniae genome. Products from the present
XX invention can be used in diagnosis kits and assays, and pharmaceutical
XX compositions and vaccines for S. pneumoniae.
```

```
XX Sequence 586 BP; 149 A; 190 C; 136 G; 111 T; 0 other;
```

PI Barash SC, Dillon PJ, Kunsch CA;
XX

OM of: US-09-301-906-5 to: Issued_Patents_NA:* out_format : pfs

Date: Feb 14, 2001 4:54 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+g2n.model -DEV=xlip
-O=/cgn2_1/USPTO.spool/US09301906/runat_31012001_152936_16831/app_query.fasta_1.271
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-GAPOP=6.000 -GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blossum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09301906_rcgni_1_52 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:

Query: US-09-301-906-5

Query length: 120

Database: Issued_Patents_NA:*

Database sequences: 280836

Search time (sec): 80580151

Search time (sec): 76.840000

score_list:

Sequence	Strd Orig	zScore	Escore	Len
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-733-837B-1	Documentation	72.50	135.44	5.44
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-638-431-1		72.50	119.21	43.61
/cgn2_6/ptodata/2/ina/5A_COMB.seq:PCT-US92-00018-1		72.50	119.21	43.61
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-821-119-1		72.00	135.77	5.21
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-821-119-16		72.00	125.18	20.26
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-555-568B-18		71.00	129.90	11.06
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-081-610-3		70.00	117.49	54.35
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-769-309A-4		70.00	110.03	141.51
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-994-570-4		70.00	110.03	141.51
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-687-289A-2		69.50	114.34	81.42
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-041-538-1		69.50	113.56	89.97
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-463-642-1		69.50	113.56	89.97
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-455-602-1		69.50	113.56	89.97
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-465-157-1		69.50	113.56	89.97
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-994-570-4		69.50	113.56	89.97
/cgn2_6/ptodata/2/ina/5B_COMB.seq:PCT-US91-09422-1		69.50	111.44	118.09
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-463-081B-30		69.00	119.26	43.32
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-461-379A-30		69.00	119.26	43.32
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-462-390B-30		69.00	119.26	43.32
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-463-074B-30		69.00	119.26	43.32
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-465-585C-30		69.00	119.26	43.32
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-652-445-30		69.00	119.26	43.32
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-472-478-1		69.00	116.54	61.38
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-463-081B-7		69.00	116.54	61.38
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-461-379A-7		69.00	116.54	61.38
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-462-390B-7		69.00	116.54	61.38
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-463-074B-7		69.00	116.54	61.38
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-465-585C-7		69.00	116.54	61.38
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-652-445-7		69.00	116.54	61.38
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-920-281C-1		69.00	101.86	403.36
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-144-602B-7		68.00	122.43	28.17
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-144-602B-7		68.00	122.34	29.17
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-651-818A-1		68.00	117.50	52.89
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-144-602B-5		68.00	117.51	54.21
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-144-602B-4		68.00	107.36	199.20
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-471-119A-1		68.00	84.56	3.7e+03
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-973-273-1		67.50	100.37	488.41
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-09-146-675-2		67.00	121.07	34.33
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-080-255-4		67.00	108.36	175.29

/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-465-713-4 + 67.00 108.36 175.29 4201 1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:PCT-US93-05857-4 + 67.00 108.36 175.29 42
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-320-559-1 + 67.00 95.20 947.36 14255
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-327-392-1 + 67.00 95.20 947.36 14255
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-306-691B-55 + 67.00 95.20 947.36 1425
seq_name: /cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-733-837B-1

seq_documentation_block:

Sequence 1, Application US/08733837B
Patent No. 6107072
GENERAL INFORMATION:
APPLICANT: Ishida, Chika
TITLE OF INVENTION: Thermostable Geranylgeranyl Diphosphate
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kenyon & Kenyon
STREET: One Broadway
CITY: New York
STATE: NY
COUNTRY: US
ZIP: 10004
COMPUTER READABLE FORM:
MEDIUM TYPE: 3+ Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
SOFTWARE: WordPerfect 6.1 Windows
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/733,837B
FILING DATE: 18-OCT-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-294956
FILING DATE: 19-OCT-1995
ATTORNEY/AGENT INFORMATION:
NAME: Greason, Edward W.
REGISTRATION NUMBER: 18,918
REFERENCE/DOCKET NUMBER: 77670/448
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-425-7200
TELEFAX: 212-425-5288
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1035 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-08-733-837B-1

alignment_scores:

Quality: 72.50 Length: 90
Ratio: 1.812 Gaps: 3
Percent Similarity: 44.44 Percent Identity: 27.778

alignment_block:

US-09-301-906-5 x US-08-733-837B-1/rev

Align seg 1/1 to reverse of: US-08-733-837B-1 from: 1 to: 1035

53 ProTyrAlaProVal..... 57
|||||
681 CCTTCTCTCCGTAGGCTCCCTCCCTCCAGGTTCAAGACGTCGTCCA 632
58LysArgGluInProLysProAlaVal..... 66
||| |||||
631 CGATCTGGAAGCGCTCCAGGCTTAAGCCCTCTCTAGCGCGC 582
67LysGlnAspGluInProLysArgGlnAlaSerHisTrp 80
|||||
581 GCGGGGCTCTCCGAGGAAGACCCCGAGGCGGACCGGC 532

```

81 AlaValLysProThrAlaValAlcGlyValHisValProLeuProLysLysGly 97
   ::::: ||||| ||||| :::::
531 GGTGTAGTAGACGGCTTGTGGCCACCATCG.....GA 457

97 nGluAlaLeuGluProAlaGlnSerValProGlnGlnSerLeuGluGluL 114
   ::::: ||||| ::::: ||| ::::: |||||
496 AGTAGTCCTCCGCGCTCAGGTCAAAGTCCACCGAGGGTCCAGAGGAGG 447

114 ysAlaAlaLeuThrPheGly 120
   ::::: ||||| |||
446 TCCAGGTGCTGACCCGTAGGC 427

```

seq_name: /cqn2_6/ptodata/2/ina/5A_COMB.seq:US-07-638-431-1

```

seq_documentation_block:
; Sequence 1, Application US/07638431
; Patent No. 5198535
;
; GENERAL INFORMATION:
;
; APPLICANT: Hoffman, Stephen L.
; APPLICANT: Charoenvit, Yupin
; APPLICANT: Hedstrom, Richard
; APPLICANT: Khusmith, Srisin
; APPLICANT: Rogers IV, William O.
;
; TITLE OF INVENTION: Protective malaria sporozoite surface protein
;
; TITLE OF INVENTION: immunogen and gene
;
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
;

```

ADDRESSEE: A. David Spevack
STREET: NMRDC Building 1 T-12 National Naval
STREET: Medical Center
CITY: Bethesda
STATE: MD

```

COUNTRY: USA
ZIP: 20814-5044
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/638,431
FILING DATE: 19910110
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Spevack, Avrom D.
TELEPHONE INFORMATION:
TELEPHONE: (301) 295-6759
TELEFAX: (301) 295-4033
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4673 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: N
ANTI-SENSE: N
ORIGINAL SOURCE:
ORGANISM: Plasmodium yoelii
STRAIN: 17X(NL)
DEVELOPMENTAL STAGE: erythrocytic
TISSUE TYPE: Blood
CELL TYPE: erythrocytic stage
IMMEDIATE SOURCE:
LIBRARY: pY-lambdagt11-2-7 kb gen
CLONE: pY10.1111
FEATURE:
NAME/KEY: CDS
LOCATION: 718..3195
OTHER INFORMATION:

```

US-07-638-431-1

```

alignment_scores:
  Quality: 72.50      Length: 70
  Ratio: 1.908        Gaps: 2
  Percent Similarity: 54.286  Percent Identity: 25.714

alignment_block:
  US-09-301-906-5 x US-07-638-431-1  ..

  Align seg 1/1  to: US-07-638-431-1  from: 1  to: 4673

51  ValValProTyrAlaProValLysArgGluGlnProLysProAla..... 65
    :::  |||  |||  |||  ::::  ::::  ||||  |||
1869  ATCAACACCCAAACACCCACCCAAAAGGCGAAACCCAAAAGCGCAACC 1918
    ::::  |||  :::  ||||  |||  :::  ||||  |||  :::  |||

66  .....ValLysGlnAspGluGlnLysProLysArgGlnAlaSerHist 80
    ::::  |||  :::  ||||  |||  :::  ||||  |||  :::  |||

1919  CAACACAGCCAAACCAACACAGCCCAACCCCAACCAAGCAACCAACCAACC 1968
    ::::  ||||  |||  :::  ||||  |||  :::  ||||  |||  :::  |||

80  rpAlaValLysProThrAlaValAlGlyValHisValProLeuProLysLys 96
    ::::  ||||  |||  :::  ||||  |||  :::  ||||  |||  :::  |||

1969  GAACCATCAACCCCAACCAAGCC.....AAACCC 1997
    ::::  ||||  |||  :::  ||||  |||  :::  ||||  |||  :::  |||

97  GlnGluAlaLeuGluProAlaGlnSerValProGlnSerLeuGluG1 113
    ::::  ||||  |||  |||  :::  ||||  |||  :::  ||||  |||  :::  |||

1998  AAACGAACCATCAACACCCCAACCAAGCCCAACCCCAACCAAGCAACCATCAACC 2047
    ::::  ||||  |||  |||  :::  ||||  |||  :::  ||||  |||  :::  |||

113  uLysAlaAla 116
    |||  |||
2048  CAACCAAGCC 2057

```

seq_name: /cqn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US92-00018-1

```

seq_documentation_block:
: Sequence 1, Application PC/TUS92000018
: GENERAL INFORMATION:
: APPLICANT: Hoffman, Stephen L.
: APPLICANT: Charcoenvit, Yupin
: APPLICANT: Hedstrom, Richard
: APPLICANT: Khushmith, Srsin
: APPLICANT: Rogers IV, William O.
: TITLE OF INVENTION: Protective ma
: TITLE OF INVENTION: immunogen and
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: A. David Spevack
: STREET: NMRDC Building 1 T-12 N
: STREET: Medical Center
: CITY: Bethesda
: STATE: MD
: COUNTRY: USA
: ZIP: 20814-5044
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.1.2
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US92/000
: FILING DATE: 19920103
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: Spevack, Avram D.
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (301) 295-6759
: TELEFAX: (301) 295-4033
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4673 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)

```



```
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/821,119
; FILING DATE: 19-MAR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4107.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2424 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-821-119-16

alignment_scores:
    Quality: 72.00      Length: 93
    Ratio: 1.358        Gaps: 4
Percent Similarity: 56.989 Percent Identity: 26.882

alignment_block:
US-09-301-906-5 x US-08-821-119-16 ..
Align seg 1/1 to: US-08-821-119-16 from: 1 to: 2424
9 ArgGlyAspGlyLysValLeuMetGluAlaArgThrSerThrSerAl 25
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
997 AGAGTTTCAGACGACCTGCTCCCAATGCCATGCCAGCAACTTCAGC 1046
25 aThrSerAspValSerAspPheAspValPheGluAlaValSerAsnA 42
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1047 GTCGTCCAAT.....TCACGG 1063
42 laLeuLeuValValHisTyrHisArgValValProTyrAlaProVal 58
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1064 CGGCTCAACATCAGCTTTCATCGAGTG.....ACAGCG 1098
58 sArgGluGlnProLysProAlaValLysGlnAspGluGln.....LysP 73
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1099 CGGAAGCAACCTCGACCTGCAGTACATCTGGGCGTCACGCGCTCCGTC 1148
73 roLysArgGlnAlaSerHisTrpAlaValLysProThrAlaValGlyVal 89
|| :|||||:|||||:|||||:|||||:|||||:|||||:
1149 CCATCACCGAGTACAGCAGCGGCGGACGAGTACGCTCCCGACCT 1198
90 HisValProLeuProLysLysGlnGlu 98
|||||:|||||:|||||:|||||:|||||:|||||:
1199 GAGCTCCCGACCCCAACGACACAG 1225

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-555-568B-18

seq_documentation_block:
; Sequence 18, Application US/08555568B
; Patent No. 5976854
; GENERAL INFORMATION:
; APPLICANT: Jones, Simon
; APPLICANT: Tang, Jim
; TITLE OF INVENTION: Calcium Independent Phospholipase A2/B
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
```

```
;
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/555,568B
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1277 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 396..1271
; US-08-555-568B-18

alignment_scores:
    Quality: 71.00      Length: 79
    Ratio: 1.732        Gaps: 4
Percent Similarity: 51.899 Percent Identity: 27.848

alignment_block:
US-09-301-906-5 x US-08-555-568B-18/rev ..
Align seg 1/1 to reverse of: US-08-555-568B-18 from: 1 to: 1277
39 ValSerAsnAlaLeuValHisTyrHisArgValValProTyrAl 55
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
322 ATCAAACTGGGCTGCTCCGCCACGCGCCCTCATCCCCCACT 273
55 aProValLysArgGluGlnProLysProAlaValLysGlnAspGluGlnL 72
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
272 CCCTCTGTCA...TCTGTCTCTTCCCA.....CAGGAAGGACTCT 235
72 ysProLysArgGlnAlaSerHis.....
|||||:|||||:|||||:|||||:|||||:|||||:
234 GTCCAGGAGAAAGCTGACGACGACGACGAGAGAAATGGGCTGGAG 185
80 TrpAlaValLysProThrAlaValGlyValHisValProLeuProLysLy 96
||| :|||||:|||||:|||||:|||||:|||||:|||||:
184 TGGAAAGGGGGGCCCTGCTCTGGGA.....
96 sGlnGluAlaLeuGluProAlaGlnSerValProGln 108
|||||:|||||:|||||:|||||:|||||:|||||:
157 ....AGGGCTCCCACTGCTCCCGACGCGCTCCAGG 125

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-081-610-3

seq_documentation_block:
; Sequence 3, Application US/08081610
; Patent No. 5445941
; GENERAL INFORMATION:
; APPLICANT: Yang, Na N
; TITLE OF INVENTION: Materials and Methods for Screening
; ADDRESSEE: Anti-Osteoporosis or Serum Lipid Lowering Agents
; STREET:
; NUMBER OF SEQUENCES: 9
```

```
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Allegretti and Witcoff, Ltd.
;; STREET: 10 S. Wacker Dr.
;; CITY: Chicago
;; STATE: IL
;; COUNTRY: U.S.A
;; ZIP: 60606
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/081,610
;; FILING DATE:
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Heaphy, Barbara A
;; REGISTRATION NUMBER: 34,619
;; REFERENCE/DOCKET NUMBER: 93,402
;; TELEPHONE: 312-715-1000
;; TELEFAX: 312-715-1234
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 3303 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; FEATURE:
;; NAME/KEY: mRNA
;; LOCATION: 2170..3303
;; FEATURE:
;; NAME/KEY: mRNA
;; LOCATION: 2214..3303
;; FEATURE:
;; NAME/KEY: mRNA
;; LOCATION: 2219..3303
;; FEATURE:
;; NAME/KEY: misc_RNA
;; LOCATION: 3301..3303
;; OTHER INFORMATION: /note= "CDS Start, codon start = 1,
;; OTHER INFORMATION: translation M"
;; FEATURE:
;; NAME/KEY: TATA_signal
;; LOCATION: 2170..2176
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: 1896..2306
;; OTHER INFORMATION: /note= "pB-301 -301 to +110"
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: 1976..2306
;; OTHER INFORMATION: /note= "pB-221 -221 to +110"
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: 2106..2306
;; OTHER INFORMATION: /note= "pB-91 -91 to +110"
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: 2137..2306
;; OTHER INFORMATION: /note= "pB-60 -60 to +110"
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: 2150..2306
;; OTHER INFORMATION: /note= "pB-47 -47 to +110"
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: 2159..2306
;; OTHER INFORMATION: /note= "pB-38 -38 to +110"
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: 2159..2271
;; OTHER INFORMATION: /note= "TGFB-3 position -38 to
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: 2159..2231
;; OTHER INFORMATION: /note= "TGFB-3 position -38 to
;; US-08-081-610-3
alignment_scores:
Quality: 70.00 Length: 80
Ratio: 1.667 Gaps: 2
Percent Similarity: 52.500 Percent Identity: 30.000
alignment_block:
US-09-301-906-5 x US-08-081-610-3/rev ..
Align seg 1/1 to reverse of: US-08-081-610-3 from: 1 to: 3303
18 GluAlaArgThrSerThrSerAlaThrSerAspValSerAspPheAspVa 34
:::|||||::: :::: ||| ||| ::::|||
2069 AAAGCCGAGCCCTTTGAAAAATACTTTGGAGTCTCTCTGTTGACGT 2020
34 lValPheGluAlaVal.....SerAsnAlaLeuLeuValV 46
||||| ||| ::::: :::::
2019 GGTAGCTGAACCTTTTTCCTCTTGGAAATCACTCTGCCACACGTGG 1970
46 alHisTyrHisArgValValProTyrAlaProValLysArgGluGlnPro 62
|| :::::|||||::: |||
1969 TGTTTTTCGAAAGTTACTCCGATGAGCCCTCCCTCCCTCCCTCCCA 1920
63 LysProAlaValLysGlnAspGluGlnLysProLysArgGlnAlaSerH1 79
||||| ||| ::::: |||
1919 GCCCGGGG.....CCCTCTGCCACGTGTCTAGT 1891
79 sTrpAlaValLysProThrAlaValGlyValHisValPro 92
|||||::: ||| ::::: |||
1890 CCGCGCGCTCGGCGCGTCTACTCTCTCTCTCTCTCTCTCTCTCTCT 1851
seq_name: /cqn2_6/ptodata/2/lna/5A_COMB.seq:US-08-769-309A-4
seq_documentation_block:
; Sequence 4, Application US/08769309A
; Patent No. 5741890
; GENERAL INFORMATION:
; APPLICANT: Scott, John D.,
; APPLICANT: Nauer, Brian J.,
; APPLICANT: Klauck, Theresa M.
; TITLE OF INVENTION: Protein Binding Domains of Gravin
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower/233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/769,309A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5741890and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/33451
; TELECOMMUNICATION INFORMATION:
```

```
;
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6605 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 192..5531
US-08-769-309A-4

alignment_scores:
    Quality: 70.00      Length: 58
    Ratio: 1.892       Gaps: 0
    Percent Similarity: 63.793      Percent Identity: 29.310

alignment_block:
US-09-301-906-5 x US-08-769-309A-4 ..
Align seg 1/1 to: US-08-769-309A-4 from: 1 to: 6605

55 AlaProValLysArgGluGlnProLysProAlaValLysGlnAspGluG1 71
   ||| |||:||||:|||||:||||:||||:||||:||||:||||:||||:
861 GCAGCATCCNAGAGAGCGAACCCCAACCAATCTACAGAGAAACCCGAAGA 910
   ||| |||:||||:|||||:||||:||||:||||:||||:||||:
71 nLysProLysArgGlnAlaSerHisTrpAlaValLysProThrAlaValG 88
   ||| |||:||||:|||||:||||:||||:||||:||||:||||:
911 GACCCTGAAGCGTGAGCAAGCCACGAGAAATTTCTCCCGAGCCGAAT 960
   ||| |||:||||:|||||:||||:||||:||||:||||:||||:
88 lyValHisValProLeuProLysLysGlnGluAlaLeuGluProAlaGln 104
   ||| |||:||||:|||||:||||:||||:||||:||||:||||:
961 CTGCCAAGCAGGTGGAGGAATGCAAGAGGAGGAGGAGAAACAAGAA 1010
   ||| |||:||||:|||||:||||:||||:||||:||||:||||:
105 SerValProGlnGlnSerLeuGlu 112
   ||| |||:||||:|||||:||||:||||:||||:||||:||||:
1011 AAAGAACCTAGCAAGTCTGCAGAA 1034

seq_name: /cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-994-570-4

seq_documentation_block:
; Sequence 4, Application US/08994570
; Patent No. 6090929
; GENERAL INFORMATION:
; APPLICANT: Scott, John D.,
; APPLICANT: Nauert, Brian J.,
; APPLICANT: Klauck, Theresa M.
; TITLE OF INVENTION: Protein Binding Domains of Gravin
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower/233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/994,570
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6090929and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/33451

;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6605 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 192..5531
US-08-994-570-4

alignment_scores:
    Quality: 70.00      Length: 58
    Ratio: 1.892       Gaps: 0
    Percent Similarity: 63.793      Percent Identity: 29.310

alignment_block:
US-09-301-906-5 x US-08-994-570-4 ..
Align seg 1/1 to: US-08-994-570-4 from: 1 to: 6605

55 AlaProValLysArgGluGlnProLysProAlaValLysGlnAspGluG1 71
   ||| |||:||||:|||||:||||:||||:||||:||||:||||:
861 GCAGCATCCNAGAGAGCGAACCCCAACCAATCTACAGAGAAACCCGAAGA 910
   ||| |||:||||:|||||:||||:||||:||||:||||:||||:
71 nLysProLysArgGlnAlaSerHisTrpAlaValLysProThrAlaValG 88
   ||| |||:||||:|||||:||||:||||:||||:||||:||||:
911 GACCCTGAAGCGTGAGCAAGCCACGAGAAATTTCTCCCGAGCCGAAT 960
   ||| |||:||||:|||||:||||:||||:||||:||||:||||:
88 lyValHisValProLeuProLysLysGlnGluAlaLeuGluProAlaGln 104
   ||| |||:||||:|||||:||||:||||:||||:||||:||||:
961 CTGCCAAGCAGGTGGAGGAATGCAAGAGGAGGAGGAGAAACAAGAA 1010
   ||| |||:||||:|||||:||||:||||:||||:||||:||||:
105 SerValProGlnGlnSerLeuGlu 112
   ||| |||:||||:|||||:||||:||||:||||:||||:||||:
1011 AAAGAACCTAGCAAGTCTGCAGAA 1034

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-687-389A-2

seq_documentation_block:
; Sequence 2, Application US/08687289A
; Patent No. 5981195
; GENERAL INFORMATION:
; APPLICANT: Fuller, Forrest H.
; APPLICANT: Krapcho, Karen J.
; APPLICANT: Hammerland, Lance G.
; TITLE OF INVENTION: CHIMERIC RECEPTORS AND METHODS FOR
; IDENTIFYING COMPOUNDS ACTIVE AT
; TITLE OF INVENTION: METABOTROPIC GLUTAMATE RECEPTORS AND
; TITLE OF INVENTION: THE USE OF SUCH COMPOUNDS IN THE
; TREATMENT OF NEUROLOGICAL DISORDERS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: FastSeq for Windows 2.0
```

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,289A
; FILING DATE: JULY 25, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/001,526
; FILING DATE: JULY 26, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 220/004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4000 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-687-289A-2

alignment_scores:
    Quality: 69.50      Length: 105
    Ratio: 1.264       Gaps: 6
    Percent Similarity: 52.381      Percent Identity: 32.381

alignment_block:
US-09-301-906-5 x US-08-687-289A-2

Align seg 1/1 to: US-08-687-289A-2 from: 1 to: 4000

9 ArgGlyAspGlyLysLysValLeuMetGluAlaArgThrSerThrSerAl 25
:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2912 CAAGGCTCTGGCAGAGC...CTGACCTTTTCAGATGCCAGCACCAGAC 2958

25 aThrSerAspValSerAspPheAsp.....ValValPheGluA 38
:||||:||||:||||:||||:||||:||||:||||:||||:
2959 CCTTTACAATGTGGAAGAGGACATACCCCTTCTGCTCAGCTTCAGCC 3008

38 laValSerAsnAlaLeuValHisArgValValProTyr 54
||||:||||:||||:||||:||||:||||:||||:||||:
3009 CTCCACGACGCCCTTCTATGTTGGTG.....CACCGACGGGGCCA... 3049

55 AlaProValLysArgGluGlnProLysProAlaValLysGlnAspGluG1 71
||||| ||| ||| ||| ||| ||| ||| ||| |||
3050 ...CCCGTGGCCACCACACACCTCTGCCACCCCTCTGACCGCAGAGA 3096

71 nLysProLysArgGlnAlaSerHisTrpAlaValLysProThrAlaValG 88
: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3097 GACCCCTCTCTCTCTCCGAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 3134

88 lyValHisValProLysGlnGluAlaLeuGluProAlaGln 104
||||:||||:||||:||||:||||:||||:||||:||||:
3135 GCTTGGCTCTCTCTCTCCGAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 3184

105 SerValProGlnGln 109
: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3185 CAGCCCGCCGAGCAG 3199

seq_name: /cgn2.6/ptodata/2/ina/5A_COMB.seq:US-08-041-538-1

seq_documentation_block:
; Sequence 1, Application US/08041538
; Patent No. 5365831
;
; GENERAL INFORMATION:
; APPLICANT: Mulvihill, Eileen R
; APPLICANT: Hagen, Frederick S
; APPLICANT: Houamed, Khaled M
; APPLICANT: Almers, Wolfhard
;
; TITLE OF INVENTION: G PROTEIN COUPLED GLUTAMATE
```

```
; TITLE OF INVENTION: RECEPTORS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend
; STREET: Stewart Street Tower, One Market Plaza
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/041,538
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/648,481
; FILING DATE:
; APPLICATION NUMBER: US 07/626,806
; FILING DATE: 12-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Parmelee, Steven W
; REGISTRATION NUMBER: 31,990
; REFERENCE/DOCKET NUMBER: 13952-6-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-467-9600
; TELEFAX: 206-623-6793
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4300 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Rattus norvegicus
; TISSUE TYPE: Cerebellum
; IMMEDIATE SOURCE:
; CLONE: 45-A
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 377..3973
; OTHER INFORMATION:
;
US-08-041-538-1

alignment_scores:
    Quality: 69.50      Length: 105
    Ratio: 1.264       Gaps: 6
    Percent Similarity: 52.381      Percent Identity: 32.381

alignment_block:
US-09-301-906-5 x US-08-041-538-1

Align seg 1/1 to: US-08-041-538-1 from: 1 to: 4300

9 ArgGlyAspGlyLysLysValLeuMetGluAlaArgThrSerThrSerAl 25
:||||:||||:||||:||||:||||:||||:||||:||||:||||:
3188 CAAGGCTCTGGCAGAGC...CTGACCTTTTCAGATGCCAGCACCAGAC 3234

25 aThrSerAspValSerAspPheAsp.....ValValPheGluA 38
: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3235 CCTTTACAATGTGGAAGAGGACATACCCCTTCTGCTCAGCTTCAGCC 3284

38 laValSerAsnAlaLeuValHisArgValValProTyr 54
||||:||||:||||:||||:||||:||||:||||:||||:
3285 CTCCACGACGCCCTTCTATGTTGGTG.....CACCGACGGGGCCA... 3325

55 AlaProValLysArgGluGlnProLysProAlaValLysGlnAspGluG1 71
||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

[illegible]

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-463-642-1

seq_documentation_block:

```

1 Sequence 1, Application US/08463642
2 Patent No. 5721107
3
4 GENERAL INFORMATION:
5
6 APPLICANT: Mulvihill, Eileen R
7 APPLICANT: Hagen, Frederick S
8 APPLICANT: Houamed, Khaled M
9 APPLICANT: Almers, Wolfhard
10 TITLE OF INVENTION: G PROTEIN COUPLED GLUTAMATE
11 TITLE OF INVENTION: RECEPTORS
12 NUMBER OF SEQUENCES: 15
13
14 CORRESPONDENCE ADDRESS:
15 ADDRESSEE: Townsend and Townsend
16 STREET: Steuart Street Tower, One Market Plaza
17 CITY: San Francisco
18 STATE: CA
19 COUNTRY: USA
20 ZIP: 94105
21
22 COMPUTER READABLE FORM:
23 MEDIUM TYPE: Floppy disk
24 COMPUTER: IBM PC compatible
25 OPERATING SYSTEM: PC-DOS/MS-DOS
26 SOFTWARE: PatentIn Release #1.24
27
28 CURRENT APPLICATION DATA:
29 APPLICATION NUMBER: US/08/463,642
30 FILING DATE: 05-JUN-1995
31 CLASSIFICATION: 435
32
33 PRIOR APPLICATION DATA:
34 APPLICATION NUMBER: US 07/672,007
35 FILING DATE: 18-MAR-1991
36 APPLICATION NUMBER: US 07/648,481
37 FILING DATE: 30-JAN-1991
38 APPLICATION NUMBER: 07/626,806
39 FILING DATE: 12-DEC-1990
40 ATTORNEY/AGENT INFORMATION:
41 NAME: Parmelee, Steven W
42 REGISTRATION NUMBER: 31,990
43 REFERENCE/DOCKET NUMBER: 13952-6-1-2
44 TELECOMMUNICATION INFORMATION:
45 TELEPHONE: 206-467-9600
46 TELEFAX: 206-623-6793
47
48 INFORMATION FOR SEQ ID NO: 1:
49 SEQUENCE CHARACTERISTICS:
50 LENGTH: 4300 base pairs
51 TYPE: nucleic acid
52 STRANDEDNESS: single
53 TOPOLOGY: linear
54 MOLECULE TYPE: cDNA
55 ORIGINAL SOURCE:
56 ORGANISM: Rattus norvegicus
57 TISSUE TYPE: Cerebellum
58 IMMEDIATE SOURCE:
59 CLONE: 45-A
60 FEATURE:
61 NAME/KEY: CDS
62 LOCATION: 377..3973
63 OTHER INFORMATION:

```

[illegible]

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq;seq:US-08-455-602-1

seq_documentation_block:

```

: Sequence 1, Application US/08455602
: Patent No. 5747267
: GENERAL INFORMATION:
: APPLICANT: Mulvihill, Eileen R
: APPLICANT: Hagen, Frederick S
: APPLICANT: Houamed, Khaled M
: APPLICANT: Almers, Wolfhard
: TITLE OF INVENTION: G PROTEIN COUPLED GLUTAMATE
: TITLE OF INVENTION: RECEPTORS
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend
: STREET: Stewart Street Tower, One Market Plaza
: CITY: San Francisco
: STATE: CA
: COUNTRY: USA
: ZIP: 94105
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.24
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/455,602
: FILING DATE: 31-MAY-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/672,007
: FILING DATE: 18-MAR-1991

```



```
; APPLICATION NUMBER: US 07/648,481
; FILING DATE: 30-JAN-1991
; APPLICATION NUMBER: 07/626,806
; FILING DATE: 12-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Parmelee, Steven W
; REGISTRATION NUMBER: 31,990
; REFERENCE/DOCKET NUMBER: 13952-6-1-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-467-9600
; TELEFAX: 206-623-6793
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4300 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Rattus norvegicus
; TISSUE TYPE: Cerebellum
; IMMEDIATE SOURCE:
; CLONE: 45-A
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 377..3973
; OTHER INFORMATION:
;
; US-08-455-602-1

alignment_scores:
  Quality: 69.50 Length: 105
  Ratio: 1.264 Gaps: 6
  Percent Similarity: 52.381 Percent Identity: 32.381

alignment_block:
  US-09-301-906-5 x US-08-455-602-1
  Align seg 1/1 to: US-08-455-602-1 from: 1 to: 4300

  9 ArgGlyAspGlyLysValLeuMetGluAlaArgThrSerAl 25
  3188 CAAGGCTCTGCGAGAGC...CTGACCTTTTCAGATGCCACCAAGAC 3234
  25 aThrSerAspValSerAspPheGlu 38
  3235 CCTTACAATGTGGAAGAGGACAAATACCCCTTCTGCTCCTCAGCC 3284
  38 laValSerAsnAlaLeuValValHisTyrHisArgValValProTyr 54
  3285 CTCCAGCAGCCCTTCTGTGTGTG.....CACCGACGGGCCA... 3325
  55 AlaProValLysArgGluGlnProLysProAlaValLysGlnAspGlu 71
  3326 ...CCCGTGGCCACCAACCACTCTGCGCCACCCCATCTGACCGGAGAGA 3372
  71 nLysProLysArgGlnAlaSerHisTrpAlaValLysProThrAlaVal 88
  3373 GACCCCTCTGTTCTCGGTGAT.....TCGGTCATCCCAAG.....G 3410
  88 lyValHisValProLeuProLysGlnGluAlaLeuGluProAlaGln 104
  3411 GCTTGCCTCCCTCTCTCCGAGCAGGACGACGACGACGACGACGACG 3460
  105 SerValProGlnGln 109
  3461 CAGCCCGGACGACGACGACGACGACGACGACGACGACGACGACGACG 3475

seq_name: /cgn2_6/plodata/2/lna/5B_COMB.seq:US-08-465-157-1
seq_documentation_block:
; Sequence 1, Application US/08465157
; Patent No. 5869609

; GENERAL INFORMATION:
; APPLICANT: Mulvihill, Eileen R
; APPLICANT: Hagen, Frederick S
; APPLICANT: Houamed, Khalid M
; APPLICANT: Almers, Wolfhard
; TITLE OF INVENTION: G PROTEIN COUPLED GLUTAMATE
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend
; STREET: Steuart Street Tower, One Market Plaza
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,157
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/041,538
; FILING DATE:
; APPLICATION NUMBER: US/07/648,481
; FILING DATE:
; APPLICATION NUMBER: US 07/626,806
; FILING DATE: 12-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Parmelee, Steven W
; REGISTRATION NUMBER: 31,990
; REFERENCE/DOCKET NUMBER: 13952-6-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-467-9600
; TELEFAX: 206-623-6793
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4300 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Rattus norvegicus
; TISSUE TYPE: Cerebellum
; IMMEDIATE SOURCE:
; CLONE: 45-A
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 377..3973
; OTHER INFORMATION:
;
; US-08-465-157-1

alignment_scores:
  Quality: 69.50 Length: 105
  Ratio: 1.264 Gaps: 6
  Percent Similarity: 52.381 Percent Identity: 32.381

alignment_block:
  US-09-301-906-5 x US-08-465-157-1
  Align seg 1/1 to: US-08-465-157-1 from: 1 to: 4300

  9 ArgGlyAspGlyLysValLeuMetGluAlaArgThrSerAl 25
  3188 CAAGGCTCTGCGAGAGC...CTGACCTTTTCAGATGCCACCAAGAC 3234
  25 aThrSerAspValSerAspPheGlu 38
  3235 CCTTACAATGTGGAAGAGGACAAATACCCCTTCTGCTCCTCAGCC 3284
  38 laValSerAsnAlaLeuValValHisTyrHisArgValValProTyr 54
  3285 CTCCAGCAGCCCTTCTGTGTGTG.....CACCGACGGGCCA... 3325
  55 AlaProValLysArgGluGlnProLysProAlaValLysGlnAspGlu 71
  3326 ...CCCGTGGCCACCAACCACTCTGCGCCACCCCATCTGACCGGAGAGA 3372
  71 nLysProLysArgGlnAlaSerHisTrpAlaValLysProThrAlaVal 88
  3373 GACCCCTCTGTTCTCGGTGAT.....TCGGTCATCCCAAG.....G 3410
  88 lyValHisValProLeuProLysGlnGluAlaLeuGluProAlaGln 104
  3411 GCTTGCCTCCCTCTCTCCGAGCAGGACGACGACGACGACGACGACG 3460
  105 SerValProGlnGln 109
  3461 CAGCCCGGACGACGACGACGACGACGACGACGACGACGACGACGACG 3475

seq_name: /cgn2_6/plodata/2/lna/5B_COMB.seq:US-08-465-157-1
seq_documentation_block:
; Sequence 1, Application US/08465157
; Patent No. 5869609
```

```
3235 CCTTTACAATGTGGAGAGAGAGACAATACCCCTTCTGCTACTTCAGCC 3284
;
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 377...3973
; PCT-US91-09422-1
;
alignment_scores:
  Quality: 69.50 Length: 105
  Ratio: 1.264 Gaps: 6
  Percent Similarity: 52.381 Percent Identity: 32.381
alignment_block:
  US-09-301-906-5 x PCT-US91-09422-1 ..
  Align seg 1/1 to: PCT-US91-09422-1 from: 1 to: 4300
;
9 ArgGlyAspGlyLysLysValLeuMetGluAlaArgThrSerAl 25
  :::::::::::::: :: ::::::::::::::
3188 CAGGCTCTGGCAAGAGC...CTGACCTTTTCAGATGCCAGCAAGAC 3234
;
25 aThrSerAspValSerAspPheasp.....ValValPheGluA 38
  :::::::::::::: ::::::::::::::
3235 CCTTTACAATGTGGAGAGAGAGACAATACCCCTTCTGCTACTTCAGCC 3284
;
38 laValSerAsnAlaLeuValHisValHisArgValValProTyr 54
  :::::::::::::: ::::::::::::::
3285 CTCCCAGCAGCCCTTCTATGGTGGTG.....CACCGACGGGGCCA... 3325
  :::::::::::::: ::::::::::::::
55 AlaProValLysArgGluGlnProLysProAlaValLysGlnAspGlu 71
  :::::::::::::: ::::::::::::::
3326 ...CCCGTGGCCACACACACCTCTGCCACCCCATCTGACCGCAGAAGA 3372
  :::::::::::::: ::::::::::::::
71 nLysProLysArgGlnAlaSerHisTrpAlaValLysProThrAlaValG 88
  :::::::::::::: ::::::::::::::
3373 GACCCCTCTGCTGCTGAT.....TCCGTATCCCAAG.....G 3410
  :::::::::::::: ::::::::::::::
88 lyValHisValProLeuProLysLysGlnGluAlaLeuGluProAlaGln 104
  :::::::::::::: ::::::::::::::
3411 GCTTGCCTCTCTCTCCCGCAGCAGCAGCCACAGCCGCCCTCAG 3460
  :::::::::::::: ::::::::::::::
105 SerValProGlnGln 109
  ::::::::::::::
3461 CAGCCCGCCGAGCAG 3475
```

seq_name: /cqn2.6/ptodata/2/ina/PCTUS_COMB.seq: PCT-US91-09422-1

```
seq_documentation_block:
; Sequence 1, Application PC/TUS9109422
; GENERAL INFORMATION:
; APPLICANT: Mulvihill, Eileen R.
; APPLICANT: Hagen, Frederick S.
; APPLICANT: Houamed, Khaled M.
; APPLICANT: Almers, Wolfhard
; TITLE OF INVENTION: G PROTEIN-COUPLED GLUTAMATE RECEPTORS
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend
; STREET: One Market Plaza, Steuart Street Tower
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105-1492
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/09422
; FILING DATE: 19911212
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/672,007
; FILING DATE: 18-MAR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/648,481
; FILING DATE: 30-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/626,806
; FILING DATE: 12-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Parmelee, Steven W.
; REGISTRATION NUMBER: 31,990
; REFERENCE/DOCKET NUMBER: 13952-6PC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 467-9600
; TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4300 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; IMMEDIATE SOURCE:
```



```

76 nAlaSerHisTrpAlaValLysProThrAlaValGlyValHisValProL 93
   ::::: ||| ||||| ||| ::|
140 AAGTACCAAAAGTTGCAGAGAAACCAAGGCCACCTCCAC 189
   ::::: ||| ||||| ||| ::|
93 euProLysLysGlnGluAlaLeuGluProAlaGlnSerValProGlnGln 109
   ||:::||:::||:::|| ||||| :::::||| ::|
190 CGCCATCCAAACAGTCACTAAAGACCG.....CAGCTTCCACCTAAG 233
   ::|
110 SerLeuGluGluLysAlaAlaLeuThr 118
   ::| ||:::||:::||:::|| :::::|||
234 CATAGGGAAGACCGGGTCTCTATGACA 260

seq_name: gb_est83:N65855

seq_documentation_block:
LOCUS N65855 533 bp mRNA EST 05-JAN-1998
DEFINITION 20895 Lambda-PRL2 Arabidopsis thaliana cDNA clone 240E217, mRNA
sequence.
ACCESSION N65855
VERSION N65855
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsi.
REFERENCE 1 (bases 1 to 533)
AUTHORS Newman,T., deBruijn,F.J., Green,P., Keegstra,K., Kende,H., McIntosh
,J., Ohlrogge,J., Raikhel,N., Somerville,S., Thomashow,M., Retzel
,E. and Somerville,C.
TITLE Genes galore: a summary of methods for accessing results from
large-scale partial sequencing of anonymous Arabidopsis cDNA clones
J Plant Physiol. 106, 1241-1255 (1994)
MEDLINE 95148729
COMMENT Contact: Thomas Newman
MSU-DOE Plant Research Laboratory
Michigan State University
MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.
Lansing, MI
Tel: 517-353-0854
Fax: 517-353-9168
Email: 22313tch@bm.cl.msu.edu
Seq primer: T7 dye primer.
FEATURES
source
1..533
Location/Qualifiers
/organism="Arabidopsis thaliana"
/strain="var columbia"
/db_xref="taxon:3702"
/clone_lib="Lambda-PRL2"
/note="Vector: Lambda Zip-Lox; Site_1: Sal; Site_2: Not;
Lambda PRL2 is a cDNA library derived from equal
quantities of 4 pools of mRNA. The mRNA sources were 1) 7
day germinated etiolated seedlings; 2) tissue culture
grown roots; 3) staged plants half with 24 hour light
cycle, half on 16 hr light, 8 hour dark- rosettes; 4)
same plants as 3 but aerial tissue (stems, flowers and
siliques. The vector is BRL's lambda Zip-Lox. The cDNA
inserts were directionally cloned with Sal-Not arms using
oligo dt primed cDNA."
BASE COUNT 140 a 119 c 127 g 130 t 17 others
ORIGIN

alignment_scores:
Quality: 88.00 Length: 109
Ratio: 1.517 Gaps: 5
Percent Similarity: 53.211 Percent Identity: 30.275

alignment_block:
US-09-301-906-5 x N65855
Align seg 1/1 to: N65855 from: 1 to: 533

```

```

12 GlyLysLysValLeuMetGluAlaArgThrSerThrSerAlaThrSerAs 28
   ||:::||:::||:::|| ||||| :::::||| ::|
18 GGAACCAAGGTT.....GCAAGANTTTNTACATCTNCG..... 50
   ::|
28 pValSerAspPheAspValValPheGluAlaValSerAsnAlaLeuLeuV 45
   ::||:::||:::||:::|| ||||| :::::||| ::|
51 .....GATGCTGTGTCTCATGTG..... 68
   ::|
45 alValHisTyrHisArgValValProTyrAlaProValLysArgGluGln 61
   ||||| :::::|||
69 .....GCACCCCTCAGAAAGGCACCA 89
   ::|
62 ProLysProAlaValLysGlnAsp.....GluGlnLysProLysArgG1 76
   ||||| ||| ||| :::::||:::||:::|| ::|
90 GAGAAACCTGCTCCCAAGCCTTCTCTCTCTCTGCTGAGAGGCCCAAGTTGA 139
   :::::||| ||| ||||| ::|
76 nAlaSerHisTrpAlaValLysProThrAlaValGlyValHisValProL 93
   :::::||| ||| ||||| ||| ::|
140 AAGTACCAAAAGTTGCAGAGAAACCAAGGCCACCTCCACCTCCAC 189
   :::::||| ||| ||||| ||| ::|
93 euProLysLysGlnGluAlaLeuGluProAlaGlnSerValProGlnGln 109
   ||:::||:::||:::|| ||||| :::::||| ::|
190 CGCCATCCAAACAGTCACTAAAGACCG.....CAGCTTCCACCTAAG 233
   ::|
110 SerLeuGluGluLysAlaAlaLeuThr 118
   ::| ||:::||:::||:::|| :::::|||
234 CATAGGGAAGACCGGGTCTCTATGACA 260

seq_name: gb_gss26:CNS0224A

seq_documentation_block:
LOCUS CNS0224A 893 bp DNA GSS 15-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone
181N04 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL220483
VERSION AL220483.1 GI:7879302
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Eurypterygii; Ctenosquamata; Acanthomorpha; Euacanthomorpha;
Holacanthopterygii; Acanthopterygii; Percomorpha;
Tetraodontiformes; Tetraodontidae; Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 893)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 893)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 893)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES
source
1..893
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="181N04"

```

```

/clone_lib="G"
/notice="Genoscope sequence ID : C0AG181DG02LP1-end : T7"
BASE COUNT 242 a 264 c 189 g 197 t 1 others
ORIGIN

alignment_scores:
  Quality: 87.00 Length: 128
  Ratio: 1.160 Gaps: 5
  Percent Similarity: 58.594 Percent Identity: 25.781

alignment_block:
US-09-301-906-5 x CNS0224A ..
Align seg 1/1 to: CNS0224A from: 1 to: 893

1 ValSerGlySerValSerAlaLeuArgGlyAspGlyLysLysValLeuMe 17
   : : : : : ||| ||| : : : : : ||| ||| : : : : : ||| |||
274 ATCGACCCAGTCGAGTACATCGCA.....AAATCTTAT 311

17 tGluAlaArgThrSerThrSerAlaThrSerAspValSerPheAspV 34
   ||| : : : : : : : : : : : : : : : : : : : : : : : :
312 GGAGCCCAATATGTCTGGCATCCCAACAGTAGTCTCGACCTTCCACCA 361

34 aValPheGluAlaValSerAsnAlaLeuValValHisThrHisArg 50
   : : : : : : : : : : : : : : : : : : : : : : : :
362 CAGTGACATCGACGTGCTCAACCCCTCTCGGGTGGCGGAAATGAAAGT 411

51 ValVal.....ProTyrAlaProValLysArgGlu... 60
   : : : : : ||| ||| : : : : : ||| ||| : : : : : ||| |||
412 CTCTCTGATACCTGGGCTCAAGACCGTTTACCCAGAGGAAGATCGTC 461

61 .....GlnProLysProAlaValLysGlnAspGluG 71
   ||| ||| ||| ||| : : : : : : : : : : : : : : : :
462 TTTACACCCCAACCTGTACACACAAACCAATTTCTGCTGACAGAGTCCC 511

71 lNlysProLysArgGlnAlaSerHisTrpAlaValLysPro...ThrAla 86
   ||| : : : : : : : : : : : : : : : : : : : : : : : :
512 AGACGTGACAGAGAAGTCCCAACATACAGTATTCTCTCTACTACTCA 561

87 ValGlyValHisValProLeuProLysLysGlnGluAlaLeuGluProAl 103
   ||| : : : : : ||| ||| ||| : : : : : : : : : : : :
562 GTAATTAGTCGAATTCATTCCTTATGATACAGAG..... 597

103 aGlnSerValProGlnGlnSerLeuGluGluLys 114
   : : : : : ||| ||| : : : : : : : : : : : : : : :
598 ....GAAACTCCAGAAATTTCACTAAGCAACCGA 627

seq_name: gb_est24:A1727801

seq_documentation_block:
LOCUS A1727801 538 bp mRNA EST 11-JUN-1999
DEFINITION BNIGH19122 Six-day Cotton fiber Gossypium hirsutum cDNA 5' similar
to (AJ223803) 2-oxoglutarate dehydrogenase E2 subunit [Arabidopsis
thaliana], mRNA sequence.
ACCESSION A1727801
VERSION A1727801.1 GI:5046653
KEYWORDS EST.
SOURCE upland cotton.
ORGANISM Gossypium hirsutum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Malvales; Malvaceae; Gossypium.
REFERENCE
  1 (bases 1 to 538)
  Blewitt, M., Matz, E.C., Davy, D.F. and Burr, B.
  ESTs from developing cotton fiber
  Unpublished (1999)
  Contact: Ben Burr
  Biology Department
  Brookhaven National Laboratory
  Upton, NY 11973, USA
  Tel: 516-344-3396
  Fax: 516-344-3407

```

```

Email: burr@bnlxl.bnl.gov
Seq primer: T3 Primer.
Location/Qualifiers
source
1..538
/organism="Gossypium hirsutum"
/cultivar="Acala Maxxa"
/db_xref="taxon:3635"
/clone_lib="Six-day Cotton fiber"
/tissue_type="immature fiber"
/dev_stage="Six days post anthesis"
/lab_host="XLI-Blue"
/notes="vector; pBluescript II KS+"

BASE COUNT 163 a 113 c 126 g 135 t 1 others
ORIGIN

alignment_scores:
  Quality: 85.50 Length: 76
  Ratio: 1.819 Gaps: 4
  Percent Similarity: 61.842 Percent Identity: 35.526

alignment_block:
US-09-301-906-5 x A1727801 ..
Align seg 1/1 to: A1727801 from: 1 to: 538

51 ValValProTyrAlaProValLysArgGlu.....G1 61
   ||| : : : : : ||| ||| : : : : : ||| ||| : : : : : ||| |||
116 GTAGCAGTCTCTCCCACTGAAAGAAATCGAAAGAGCTGCTTCTGA 165

61 nProLysPro.....AlaValLysGlnAspGluGlnLysProLysArgG 76
   : : : : : ||| ||| : : : : : ||| ||| : : : : : ||| |||
166 GCATCTCTCTGCTGAAGCTGTAAAGGAGGATAGCCAAAGCTAAAGTTG 215

76 lNAlaSerHisTrpAlaValLysProThrAlaValGlyValHisValPro 92
   : : : : : ||| ||| ||| ||| : : : : : ||| ||| : : : : : ||| |||
216 AGCTTCTCTCTGCTGGCGGAGCCCTAAAGCCCTTCT.....CCA 256

93 LeuProLysLysGlnGluAlaLeuGluProAlaGlnSerValProGlnG1 109
   ||| ||| : : : : : ||| ||| ||| : : : : : ||| ||| : : : : : ||| |||
257 CCACCTCCAAACGACACAGCTTCTGAACCTGTA.....CTTCCACCCAA 300

109 nSerLeuGluGluLysAlaAlaLeuThr 118
   : : : : : ||| ||| ||| : : : : : ||| ||| : : : : : ||| |||
301 GGAAAGGGAAGAAAGGTTCTTATGACA 328

seq_name: gb_gss27:CNS04DK7

seq_documentation_block:
LOCUS CNS04DK7 888 bp DNA GSS 21-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
102B08 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL285856.1 GI:8024290
VERSION AL285856.1
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Eurypterygii; Ctenosquamata; Acanthomorpha; Eucanthomorpha;
Holacanthopterygii; Acanthopterygii; Percomorpha;
Tetraodontiformes; Tetraodontoidae; Tetraodontidae; Tetraodon.
1 (bases 1 to 888)
Roest-Crollius, H., Jaillon, O., Dasilva, C., Fizes, C., Fisher, C.,
Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and
Weissenbach, J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Unpublished
REFERENCE
  2 (bases 1 to 888)
  Roest-Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,
  Bernot, A., Fizes, C., Wincker, P., Brottier, P., Quetier, F.,

```



```

SOURCE      human.
ORGANISM    Homo sapiens
REFERENCE   1 (bases 1 to 820)
AUTHORS     NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert.Strausberg@nih.gov
            Tissue Procurement: ATCC/DCTD/DRP
            cDNA Library Preparation: Life Technologies, Inc.
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNL at:
            http://image.llnl.gov
            Plate: LICM751 row: o column: 06
            High quality sequence start: 50
            High quality sequence stop: 459.

FEATURES
  source
    1..820
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="IMAGE:325469"
    /clone_lib="NIH_MGC_72"
    /tissue_type="melanotic melanoma"
    /lab_host="DH10B (phage-resistant)"
    /notes="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
    Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
    Average insert size 2 Kb. Library constructed by Life
    Technologies."
  BASE COUNT  188 a  325 c  189 g  118 t
  ORIGIN
    188 a  325 c  189 g  118 t

alignment_scores
  Quality: 83.50      Length: 101
  Ratio: 1.606       Gaps: 3
  Percent Similarity: 51.485      Percent Identity: 27.723

alignment_block
  US-09-301-906-5 x BE98017 ..
  Align seg 1/1 to: BE98017 from: 1 to: 820

20 ArgThrSerThrSerAlaThrSerAspValSerAsp..... 31
|||||
367 AGACACGACCAACATGACCTCTGCTGAGCCGACCTGGGGTCAACC 416
32 .....PheAspValValPheGluAlaValSerAsnAlaLeuValV 46
|||||
417 ATCGATCTCATCAATCTCGACACCTACCGCATCGACGCCAATGTTCT 466
|||||
46 alHisTyHisArgValValProTyAlaProValLysArgGluGlnPro 62
|||||
467 AGCATCCAGACATGATGACACACTCAAGCCAGCAGGAGATCCAGGCC 516
|||||
63 LysProAlaValLysGlnAspGluGlnLysProLysArgGlnAlaSerHi 79
|||||
517 CCACCCAGCTCCAGCAGCAGCCAGCAAGCCGCGCGG..... 555
79 Strp.....AlaValLysProThrAlaValG 88
|||||
556 .TGTTGGCATGGCATGCCAACCACGACGATACATCTCCACTGCAGTCA 604
88 lyValHisValProLeuProLysLysGlnGluAlaLeuGluProAlaGln 104
|||||
605 CACGCCACTGTCATCTCCCATTTGGCCAGCCTGCGGTCCAGCATCTGCAG 654
105 Ser 105
|||||

```

```

655 AGT 657
seq_name: gb_est15:AI068871
seq_documentation_block: 761 bp mRNA EST 09-DEC-1999
LOCUS AI068871 mgae0004cc04f Magnaporthe grisea Appressorium Stage cDNA Library
DEFINITION Pyricularia grisea cDNA clone mgae0004cc04f 5', mRNA sequence.
ACCESSION AI068871
VERSION AI068871.1 GI:3391846
KEYWORDS EST.
SOURCE Pyricularia grisea.
ORGANISM Pyricularia grisea.
REFERENCE Magnaporthe Fungi; Ascomycota; Ascomycota incertae sedis;
AUTHORS Magnaportheaceae; anamorphic Magnaportheaceae; Pyricularia.
TITLE 1 (bases 1 to 761)
AUTHORS Choi, W., Fang, E., Sasinowski, M., Wing, R. and Dean, R. A.
TITLE Expressed sequence characterization during appressorium formation
JOURNAL in rice blast fungus, Magnaporthe grisea
COMMENT Unpublished (1998)
CONTACT: Dean, R. A.
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson University, Clemson, SC 29634, USA
Tel: 864 656 5737
Fax: 864 656 4293
Email: rdean@clemson.edu
Seq primer: T3 primer (AATTAAACCTCTCACTAAAGG)
High quality sequence stop: 261.

FEATURES
  source
    1..761
    /organism="Pyricularia grisea"
    /strain="70-15"
    /db_xref="taxon:89476"
    /clone="mgae0004cc04f"
    /clone_lib="Magnaporthe grisea Appressorium Stage cDNA
    Library"
    /dev_stage="Germinated conidia on appressorium-inductive
    surface"
    /note="Vector: pBlueScriptII SK(+) Vector; Site_1: EORI;
    Site_2: XhoI; The appressorium formation-specific cDNA
    library was constructed from conidia germinated for 5-8
    hr on an inductive surface. The library has an average
    insert size of 1.5 kbp."
  BASE COUNT  127 a  232 c  249 g  142 t  11 others
  ORIGIN
    127 a  232 c  249 g  142 t  11 others

alignment_scores
  Quality: 82.50      Length: 123
  Ratio: 1.289       Gaps: 6
  Percent Similarity: 52.033      Percent Identity: 29.268

alignment_block
  US-09-301-906-5 x AI068871/rev ..
  Align seg 1/1 to reverse of: AI068871 from: 1 to: 761

9 ArgGlyAspGlyLysLysValLeuMetGluAlaArg.....ThrSerTh 23
|||||
538 CGGGAGGTGGTCNCCGATCCAGGGGAGAGATGCTGCTCCAGATTGGAC 489
23 rSerAlaThrSerAspValSerAspPheAspValValPheGluAlaValS 40
|||||
488 CAGTCCCGCCGGACCTGGGTGAAGTAGACCCCGGGATCGCGCGCTCC 439
40 erAsnAlaLeuValValHisTyHisArgVal.....ValProTy 54
|||||
438 CGCAT.....GTCCATGACGACGTGGTGGAGGGGGCGCCAGCC 401
55 AlaProValLys.....ArgGluGlnProLysProAlaValLysGl 68
|||||
400 GCGCCCGAGACTCGAGCNCGCCAGCCAGCCGCGGATCGCGGCG 351

```

Align seg 1/1 to reverse of: BE561385 from: 1 to: 1018

```
6 SerAlaLeuArgGlyAspGlyLysValLysLysValLeuMetGluAlaAargThrSe 22  
|||||:|||||||:  
380 TCAGTGTGCAGCAGCG.....GTGGTACCCG 325  
  
85 hrAlaValGlYvalHisValProLeuPro.....94  
|||||:|||||||:  
324 TCCCACATGTCCTACCAGTGCCCGTCGAGTAGTGATGATGAC 275  
  
95 ...LysLysGlnGluAlaLeuGluProAlaGlnSerValProGlnGInse 110  
::: ||| :::::||| :|||: |  
274 GTACGCCAGCACGAGTTCCAGTCCAGCCGCCGCCAGGTCTCCGTCAACA 225  
  
110 rIcuGluGluLysAlaAla 116  
::::||| :::::|||  
224 GGTGAAGAAGGTGGCC 206
```

seq_name: gb_est72:BE561385

seq_documentation_block:

LOCUS BE561385 1018 bp mRNA EST 15-AUG-2000
DEFINITION 601344976F1 NIH_MGC_8 Homo sapiens cDNA clone IMAGE:3677634 5', mRNA sequence.

ACCESSION BE561385
VERSION BE561385.1 GI:9805105
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1018)
NTH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel.: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LICW353 row: h column: 19
High quality sequence stop: 697.
Location/Qualifiers
1..1018
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image=3677634
/clone_lib="NIH_MGC_8"
/tissue_type="Burkitt lymphoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lymph; Vector: pOMB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GCACAGC(G). Size selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong In the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 297 a 215 C 348 g 158 t

ORIGIN

alignment_scores:
Quality: 81.00 Length: 100
Ratio: 1.397 Gaps: 3
Percent Similarity: 58.000 Percent Identity: 26.000

alignment_block:
US-09-301-906-5 x BE561385/rev ..


```

US-09-301-906-5 x AI728067  ..
Align seg 1/1 to: AI728067 from: 1 to: 541

51 ValValProTyAlaProValLysArgGlu.....G1 61
|||||  |||||||.....:
330 GTACGAGTGTCTCCACTGAAAGAAATCCGAAAGGCTGTTCTGA 379
|||||  |||||||.....:
61 nProLysPro.....AlaValLysGlnAspGluGlnLysProLysArgG 76
|||||  |||||||.....:
380 GCCATCTCTGCTGAAGCTGTAAAGGAGGATAAGCCAAAGCTAAAGTTG 429

76 InAlaSerHisTrpAlaValLysProHisAlaValGlyValHisValPro 92
|||||  |||||||.....:
430 AAGTCTCTGCTGCTGCGGCGAAACCTAAAGCCCTTCT.....CCA 470

93 LeuProLysLysGlnGluAlaLeuProAlaGlnSerValProGlnG1 109
|||||  |||||||.....:
471 CCACCTCCAAACGACAGCTTCTGAACCTGTA.....CTTCCACCAA 514

109 nSerLeuGluGluLysAlaLeu 117
:|||||  |||||||.....:
515 GGAAGGAAAGAAAGGTTCTCTATG 539

seq_name: gb_gss2:AQ156493

seq_documentation_block:
LOCUS      AQ156493      599 bp      DNA      12-SEP-1998
DEFINITION nbxb0008A13f CUGI Rice BAC Library Oryza sativa genomic clone
ACCESSION  AQ156493
VERSION    AQ156493.1 GI:3553518
KEYWORDS   GSS.
SOURCE     Oryza sativa.
            Oryza sativa
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
            Wing,R.A. and Dean,R.A.
            A BAC End Sequencing Framework to Sequence the Rice Genome
            Unpublished (1998)
            Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: twing@clemson.edu
            Seq primer: TAATACGACTCACTATAGG
            Class: BAC ends
            High quality sequence start: 49
            High quality sequence stop: 318.
            Location/Qualifiers
                1..599
                /organism="Oryza sativa"
                /strain="Japonica"
                /cultivar="Nipponbare"
                /db_xref="taxon:4530"
                /clone="nbxb0008A13f"
                /clone_lib="CUGI Rice BAC Library"
                /tissue_type="Leaf"
                /lab_host="E. coli DH10B"
                /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                HindIII; Rice is one of two most popular grains in the
                world. Half of the world population especially those
                inhabiting highly populated areas of the humid tropics
                and subtropics, rely on rice as their primary source of
                carbohydrate. Monocotyledonous rice is a diploid plant
                (2n=24) with a haploid genome equivalent of 431 Mbp
                (Armuganathan and Earle, 1991). The relatively small
                genome of rice, three times larger than that of
                Arabidopsis, makes it suitable for genomic studies. In
                order to facilitate positional cloning, physical mapping
                and genome sequencing of rice, we have constructed a BAC
                library from Oryza sativa, Nipponbare variety. The
                library contains 36,864 clones with an average insert size
                of 128.5 Kb providing 10.9 haploid genome equivalents. The
                deep coverage allows the isolation a particular sequence
                with a probability of 99.9 %. Two high density filters,
                each containing 18,432 clones (doubly spotted), represent
                the whole library for colony screening."
                BASE COUNT      155 a      116 c      162 g
                ORIGIN
                alignment_scores:
                    Quality:      80.50      Length:      99
                    Ratio:        1.750      Gaps:          3
                    Percent Similarity: 46.465      Percent Identity: 26.263
                alignment_block:
                    US-09-301-906-5 x AQ156493  ..
                    Align seg 1/1 to: AQ156493 from: 1 to: 599
                    37 GluAlaValSerAsnAlaLeuValValHisValHisArgValValPr 53
                    :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
                    201 CAGATAATGACCAATGGTTGTTAAAGAGCCCGGTGCATAGGTCGTGAC 250
                    53 oTyraAla.....P 56
                    :|||
                    251 GAACGCCGAGAGGAGCGAGTGTGGTGGTGTCTTACGCGCTGGATC 300
                    56 roValLysArgGluGlnProLysProAlaValLysGlnAspGluGlnLys 72
                    :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
                    301 CGGAGAGGAGCTGCGCGCGGAGCTGGTGGAGCATGAGAGAGG 350
                    73 ProLysArgGlnAla..... 77
                    :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
                    351 CCGAGGCGAGTATGCCAGATGAAGATCAAGGACTTACTTAAGTGGCTTTTA 400
                    78 .SerHisTrpAla.....ValLysProT 85
                    :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
                    401 CGAGAACTTTCAGAGGTACACCGGCGCATCTACTCTTGTAAAGAGA 450
                    85 hrAlaValGlyValHisValProLysLysGlnGluAlaLeu 100
                    :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
                    451 CCGTATAGGATGATTGTTCCATTCTCTGTATCAGGACTTTC 497

seq_name: gb_est16:AI103843

seq_documentation_block:
LOCUS      AI103843      519 bp      mRNA      EST      31-JAN-1999
DEFINITION EST213132 Normalized rat heart, Bento Soares Rattus sp. cDNA clone
RHEBT77 3' end, mRNA sequence.
ACCESSION  AI103843
VERSION    AI103843.1 GI:3708329
KEYWORDS   EST.
SOURCE     Rattus sp.
            Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
            Rattus 1 (bases 1 to 519)
            Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
            Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
            Gene Index
            Unpublished (1998)
            Other_ESTs: TC53031
            Contact: Lee, NH
            ATCC
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208

```


/organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone_lib="RPCI-98"
 /clone="BACR14N18"
 /note="end : T7"
 BASE COUNT 301 a 230 c 208 g 185 t 45 others
 ORIGIN

alignment_scores:
 Quality: 80.00 Length: 109
 Ratio: 1.509 Gaps: 4
 Percent Similarity: 48.624 Percent Identity: 26.606
 alignment_block:
 US-09-301-906-5 x CNS00769 ..
 Align seg 1/1 to: CNS00769 from: 1 to: 969

12 GlyLysValLeuMetGluAlaArgThrSerThrSerAlaThrSerAs 28
 545 GGCACAAAAGTGACTCCGACTCCTAGTCGACTGCTCTGGGCTTCAACC 594
 28 pValSerAspPheaspValValPheGluAlaValSerAsnAlaLeuLeuV 45
 595 CAAGTCGAAGTTTCGATGAAAAGTTCGAAAAGGTA..... 628
 45 alValHisTyrHisArgValValProTyrAlaProValValLysArgGluGln 61
 629GTGAGTCACCTCAGCCGGTTKTCGAACCCCAA 661
 62 ProLysProAlaVal.....LysGlnAspGluGlnLysProLysArgG 76
 662 CCCAAGCCCGTAATCGATTTCAATTTGGATGAGTTAAACACACGTA... 709
 76 nAlaSerHisTrpAlaValLysProThrAlaValGlyValHisValProL 93
 710CCGAACCTTGAGGAGCGACCCCAAGAG.....CAGC 740
 93 euProLysLysGlnAlaLeuGluProAlaGlnInsValProGlnGln 109
 741 TGCCAGGCCTGAAGTTTGAAAACCTCCACACACAAAACCGAAGCCA 790
 110 SerLeuGluGluLysAlaLeuThr 118
 791 GCAGCACAACGTCGAGCGCTCCAACA 817

seq_name: gb_est72:BE547899

seq_documentation_block:
 LOCUS BE547899 976 bp mRNA EST 09-AUG-2000
 DEFINITION 601078151F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:3464052 5',
 mRNA sequence.
 ACCESSION BE547899
 VERSION BE547899.1 GI:9776544
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 976)
 NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: Incyte Genomics, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov
 Plate: L1AM8464 row: e column: 13
 High quality sequence start: 14
 High quality sequence stop: 286.
 FEATURES
 source 1..976
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3464052"
 /clone_lib="NIH_MGC_12"
 /tissue_type="choriocarcinoma"
 /lab_host="DH108"
 /note="Organ: placenta; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 1.4 kb. Library prepared by Life
 Technologies."

BASE COUNT 221 a 421 c 151 g 183 t
 ORIGIN

alignment_scores:
 Quality: 80.00 Length: 53
 Ratio: 2.581 Gaps: 3
 Percent Similarity: 58.491 Percent Identity: 37.736
 alignment_block:
 US-09-301-906-5 x BE547899 ..
 Align seg 1/1 to: BE547899 from: 1 to: 976
 46 ValHisTyrHisArg.ValValProTyrAlaProVal.....L 58
 606 GTGCACGCCCATTTATCGCAGTCCCTACGGCGCTACCCACTGGCGCCGCC 655
 58 ysArgGluGlnProLysProAlaValLysGlnAspGluGlnLysProLys 74
 656 AGCCCGCGCAGCCACGCCCTACCCCTG.....CCCAACCCCTCC 693
 75 ArgGlnAlaSerHisTrpAlaValLysProThrAlaValGlyValHisVa 91
 694 GCTCACTTCCCCCACTGGCCACCCGCCCCACACAGGGGGGAACCCACCG 743
 91 lProLeu 93
 744 ACCCGTG 750

This Page Blank (uspto)

OM of: US-09-301-906-15 to: EST:* out_format : pfs
Date: Feb 18, 2001 3:14 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
-MODEL=frame-p2n.model -DEV=slp
-Q/cgml1/USPTO.spool/US09301906/runat_15022001_103218_9304/app_query.fasta_1.2332
-DB=EST -QFMT=fastap -SUFFIX=fst -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -OGAPOP=4.500
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FCAPOP=6.000
-FGAPEXT=7.000 -XGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosom62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=2000000000 -USER=US09301906 -CGNL_1_6280 -NCPU=6
-ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-301-906-15
Query length: 2237
Database: EST*
Database sequences: 7991742
Database length: 791223438
Search time (sec): 3263.170000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
gb_est21:AI546664	+	235.00	424.13	1.3e-14	580	! AI546664 SD09861.5prime SD Dros
gb_est21:AI542434	+	226.00	406.35	1.3e-13	592	! AI542434 SD08761.5prime SD Dros
gb_est21:AI531540	+	171.50	299.59	1.1e-07	629	! AI531540 SD02626.5prime SD Dros
gb_est21:AI532043	+	162.00	280.97	1.2e-06	636	! AI532043 SD03401.5prime SD Dros
gb_est21:AI531508	+	159.00	279.13	1.5e-06	482	! AI531508 SD02578.5prime SD Dros
gb_est21:AI546576	+	158.50	275.27	2.5e-06	589	! AI546576 SD06393.5prime SD Dros
gb_est21:AI530886	+	158.00	274.25	2.9e-06	591	! AI530886 SD01648.5prime SD Dros
gb_est21:AI543590	+	155.00	268.70	5.9e-06	580	! AI543590 SD10387.5prime SD Dros
gb_est21:AI542293	+	149.50	258.18	2.3e-05	573	! AI542293 SD08556.5prime SD Dros
gb_est21:AI542833	+	144.00	251.04	5.7e-05	448	! AI542833 SD09311.5prime SD Dros
gb_est40:AW160301	+	139.50	238.80	0.0003	571	! AW160301 EST290159 L. pennellii
gb_est21:AI533122	+	139.00	236.95	0.0003	607	! AI533122 SD02020.5prime SD Dros
gb_est21:AI542359	+	135.00	237.28	0.0003	346	! AI542359 SD08652.5prime SD Dros
gb_est21:AI532485	+	128.50	219.69	0.0032	488	! AI532485 SD04021.5prime SD Dros
gb_est21:AI531921	+	126.00	216.69	0.0047	429	! AI531921 SD03228.5prime SD Dros
gb_est21:AI534252	+	126.00	209.64	0.0115	700	! AI534252 SD07019.5prime SD Dros
gb_est21:AI533189	+	125.50	212.17	0.0083	549	! AI533189 SD04930.5prime SD Dros
gb_est21:AI534525	+	125.50	211.23	0.0094	586	! AI534525 SD07414.5prime SD Dros
gb_est21:AI541573	+	125.50	209.98	0.0110	639	! AI541573 SD02710.5prime SD Dros
gb_est21:AI532597	+	124.50	208.18	0.0139	633	! AI532597 SD04156.5prime SD Dros
gb_est21:AI534952	+	124.50	205.83	0.0187	745	! AI534952 SD01247.5prime SD Dros
gb_est69:BE285583	+	123.00	204.73	0.0216	657	! BE285583 601098011F1 NCI CGAP_M
gb_est21:AI544309	+	122.50	206.34	0.0176	549	! AI544309 SD10815.5prime SD Dros
gb_est21:AI533624	+	122.50	203.69	0.0247	660	! AI533624 SD05519.5prime SD Dros
gb_est69:BE267444	+	120.50	195.82	0.0677	870	! BE267444 601189703F2 NIH_MGC_7
gb_est69:BE270270	+	119.50	196.51	0.0620	725	! BE270270 601186063F1 NIH_MGC_8
gb_gss28:CN50516A	+	117.00	189.41	0.1540	847	! BE741486 601594529F1 NIH_MGC_9
gb_est75:BE741486	+	116.00	184.02	0.3074	1076	! AI342379 Tetraodon nigroviridi
gb_est75:BE745641	+	115.00	185.27	0.2619	862	! BE745641 601578722F1 NIH_MGC_9
gb_est78:BE870025	+	114.00	186.96	0.2109	670	! BE870025 60149777F1 NIH_MGC_65
gb_est78:BE952252	+	112.00	180.77	0.4664	786	! BE952252 UI-M-CD0-ayk-e-05-0-UI
gb_est11:AA754585	+	111.50	184.94	0.2732	550	! AA754585 97SN1967 Rice Immature
gb_est43:AW330438	+	111.50	181.55	0.4220	696	! AW330438 PFOAFBC525SK Onchocera
gb_est10:AA695102	+	111.50	181.02	0.4515	722	! AA695102 GM02143.5prime GM Dros
gb_est68:BE216933	+	111.50	176.87	0.7687	963	! BE216933 EST0476 Triticum aestiv
gb_est77:BE880814	+	110.00	168.02	2.39	1455	! BE880814 601493243F1 NIH_MGC_6
gb_est69:BE294562	+	109.50	178.19	0.6492	671	! BE294562 601176132F1 NIH_MGC_17
gb_est75:BE778378	+	109.50	172.30	1.38	1010	! BE778378 601463788F1 NIH_MGC_6
gb_est77:BE908347	+	108.50	172.41	1.36	876	! BE908347 601503153F1 NIH_MGC_70
gb_est75:BE785467	+	108.50	168.97	2.12	1112	! BE785467 601478038F1 NIH_MGC_9
gb_est77:BE882707	+	108.00	163.61	4.21	1508	! BE882707 601511174F1 NIH_MGC_4

gb_est28:AJ396036 + 107.50 172.32 1.38 770 ! AJ396036 AJ396036 dkf42426 Ga
gb_est75:BE792757 - 107.50 169.24 2.05 954 ! BE792757 601584613F1 NIH_MGC
gb_est48:AW696578 - 106.00 159.75 6.91 1506 ! AW696578 NF108F04ST1F1042 D
gb_est77:BE891163 + 105.50 158.10 8.54 1578 ! BE891163 60143282F1 NIH_MG
seq_name: gb_est21:AI546664
seq_documentation_block: 580 bp mRNA EST 22-MAR-1999
LOCUS AI546664
DEFINITION SD09861.5prime SD Drosophila melanogaster Schneider L2 cell culture
pot2 Drosophila melanogaster cDNA clone SD09861 5prime, mRNA
Drosophila melanogaster
ACCESSION AI546664
VERSION AI546664.1 GI:4464037
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 580)
AUTHORS Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C., Brokstein
P., Lewis,S. and Rubin,G.M.
TITLE BDGP/HMI Drosophila EST Project
JOURNAL Unpublished (1997)
COMMENT Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: 98 row: F column: 1
High quality sequence stop: 461.
FEATURES
source 1..580
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="SD09861"
/clone_lib="SD Drosophila melanogaster Schneider L2 cell
culture pot2"
/lab_host="DH5-alpha"
/note="vector: pot2; Site_1: EcoRI; Site_2: XhoI; Sized
fractionated cDNAs were directly ligated into pot2.
Plasmid cDNA library."
BASE COUNT 167 a 124 c 122 g 167 t
ORIGIN
alignment_scores:
Quality: 235.00 Length: 205
Ratio: 2.043 Gaps: 7
Percent Similarity: 56.098 Percent Identity: 34.146
alignment_block:
US-09-301-906-15 x AI546664 ..
Align seg 1/1 to: AI546664 from: 1 to: 580
2006 ValLeuValAspGluValHisMetMetHisGlnGlyLeuLeuGlnLeuG1 2022
17 ATCTTTATTGACGAAGCGGTGATGTCCTCCATCCAGGAGCTGTTTTTATGC 66
2022 yValPheAlaThrGlyAlaSerGluGlyLeuPheGlyAspIleAsnG 2039
:||||: :||||: :||||: :||||: :||||: :||||: :||||: :||||: :
67 TGTTCGATGCTCTGGTCGCTCAAAAGTTAGATTCTCGGTGACAACCTGC 116
2039 InIleProPheIleAsnArgGluLysValPheArgMetAsp..... 2052
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
117 AAATTCATTGTAACCGCTGCCAGCTCAAGATGAGTACGATCGT 166
2053 CysAlaValPheValProLysLysGluSerValValTyThrSerLysSe 2069
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
167 TTGGAAAAATGTCCTCAATTTGCGTTTTCATCTTTTATC.....TC 207

```
2069 rTyrArgCysProLeuAspValCysTyrLeuLeuSerSerMetThrValA 2086
|||||
208 GTACCGCTGCTCGATGCTGTTAT.....A 236
|||||
2086 rGlyThrGluLysCysTyrProGluLysValValSerClyLysAspLys 2102
|||||
237 GGTTTTACCGACACACCTGAGTGAAGTGACAAAAAGCAGGAAAGAT... 283
|||||
2103 ProValValArgSerLeuSerLysArgPro..... 2112
|||||
284 .....GTAGGTTTCATGAGTGCAGCGCGCTGTTAATACCTGTAGTT 327
|||||
2113 .....lleGlyThrThrAspValAlaGluIleAsnAlaAspValT 2127
|||||
328 TGTCCGAATATCGACATGAATGAGTCCGCTCCGTAAGGAAGTTAAAT 377
|||||
2127 rLeuCysMetThrGlnLeuGluLysSerAspMetLysArgSerLeuLys 2143
|||||
378 ATCTCGTCTTTACACAGCCGCAAAACCGAACTTGCCATG..... 418
|||||
2144 GlyLysGlyLysGluThrProValMetThrValHisGluAlaGlnGlyLy 2160
|||||
419 .....AGAATATAATGCTCTCGACAGTACATGAATCCAGGCAA 459
|||||
2160 sThrPheSerAspValValPheArgThrLysLys...AlaAspAsps 2176
|||||
460 AGAGGCTAAACCATATACCTCGTCAGACTTAACCAATATCCACAGCGG 509
|||||
2176 rLeuPheThrLysGlnProHisIleLeuValGlyLeuSerArgHisThr 2192
|||||
510 ACGTATTCTCGACATCCGTCAGTCTTAGTCGCTTAACGTCGCATACT 559
|||||
2193 ArgSerLeuValTyr 2197
|||||
560 GAAGAACTTGCTAT 574
|||||
```

seq_name: gb_est21:A1542434

seq_documentation_block: 592 bp mRNA EST 22-MAR-1999
LOCUS A1542434
DEFINITION SD08761.5prime SD Drosophila melanogaster Schneider L2 cell culture
pot2 drosophila melanogaster cDNA clone SD08761 5prime, mRNA
sequence.

ACCESSION A1542434
VERSION A1542434.1 GI:4459807
KEYWORDS EST.
SOURCE fruit fly.

ORGANISM

Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

AUTHORS Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C., Brokstein
P., Lewis,S. and Rubin,G.M.
TITLE BDGP/HMMI Drosophila EST Project
JOURNAL Unpublished (1997)
COMMENT Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 USA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu
Plate: 87 row: F column: 1
High quality sequence stop: 538.

FEATURES

source Location/Qualifiers
1..592
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="SD08761"
/clone_lib="SD Drosophila melanogaster Schneider L2 cell
culture pot2"
/lab_host="DH5-alpha"

/note="vector: pot2; Site_1: EcoRI; Site_2: XhoI; Sized
fractionated cDNAs were directly ligated into pot2.
Plasmid cDNA library."

BASE COUNT 173 a 124 c 127 g 168 t
ORIGIN

alignment_scores:

Quality: 226.00 Length: 212
Ratio: 1.948 Gaps: 8
Percent Similarity: 54.717 Percent Identity: 33.962

alignment_block:

US-09-301-906-15 x A1542434 ..

Align seg 1/1 to: A1542434 from: 1 to: 592

```
2010 GluValHisMetMetHisGlnGlyLeuLeuGlnGlyValPheAlaTh 2026
|||||
3 GAAGCCTTGAATGTCCTCCATCCAGAGCTGTTTTTATGCTGTTGGATGTC 52
|||||
2026 rGlyAlaSerGluGlyLeuPhePheGlyAspIleAsnGlnIleProPheI 2043
|||||
53 TGGTGGCTCAAAAGTAGATTCTCGTGACAACTTGCAAAATTCATTTG 102
|||||
2043 leAsnArgGluLysValPheArgMetAsp.....CysAlaValPhe 2056
|||||
103 TAAACCGTCTCCGCGAGTTCAGATGAGTTACGATCGTTGGAAAAATG 152
|||||
2057 ValProLysLysGluSerValValTyrThrSerLysSerTyrArgCysPr 2073
|||||
153 GTCCCAATTGTCGGTTCACCTTTTATC.....TCGTACCGCTGTCC 193
|||||
2073 oLeuAspValCysTyrLeuLeuSerSerMetThrValArgGlyThrGluL 2090
|||||
194 TGTCCGATGCTGTTAT.....AGGTTTACCGAGC 222
|||||
2090 ysCysTyrProGluLysValValSerGlyLysAspLysProValValArg 2106
|||||
223 ACTACCGTGAAGTGAACAAAAAGCAGGAAAGAT.....GTAGGT 263
|||||
2107 SerLeuSerLysArgPro.....IleGl 2114
|||||
264 TTGATGAGTGCCTCGCGACGGTGTAAATACCTGAGGTTTGCCTGAATATC 313
|||||
2114 yThrThrAspAspValAlaGluIleAsnAlaAspValTyrLeuCysMetT 2131
|||||
314 GACATTGAATGAGTCCCGCTCCGTAAGGAAGTTAAATATATCTCGCTTTA 363
|||||
2131 hrGlnLeuGluLysSerAspMetLysArgSerLeuLysGlyLys 2147
|||||
364 CACAGCCGAAAAACCGAAGTTCGATG.....AAG 395
|||||
2148 GluThrProValMetThrValHisGluAlaGlnGlyLysThrPheSerAs 2164
|||||
396 AAATATAATGCTCTCGACAGTACATGAATTCACGCAAGGAGGCTAAAAA 445
|||||
2164 pValValLeuPheArgThrLysLys...AlaAspAspSerLeuPheThrL 2180
|||||
446 CATATACCTCGTCAGACTTAACCAATATCCACAGACGAGGATTTCTTTCG 495
|||||
2180 ysGlnProHisIleLeuValGlyLeuSerArgHisThrArgSerLeuVal 2196
|||||
496 GACATCCGTCAGCTCTTAGTCGGCTTTAACTCGTCATACGAAAGACTTGC 545
|||||
2197 TyrAlaAlaLeuSerSerLysLeuAspLysVal 2208
|||||
546 TAT.....TACACACGACGAGAACGGATGATCTGTG 575
|||||
```

seq_name: gb_est21:A1531540

seq_documentation_block: 629 bp mRNA EST 18-MAR-1999
LOCUS A1531540

```

DEFINITION SD02626.5prime SD Drosophila melanogaster Schneider L2 cell culture
            pOT2 Drosophila melanogaster cDNA clone SD02626 5prime, mRNA
            sequence.
ACCESSION  AI531540
VERSION     AI531540.1 GI:4445675
KEYWORDS   EST.
SOURCE     fruit fly.
ORGANISM   Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE  1 (bases 1 to 629)
AUTHORS    Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C., Brokstein
            ,P., Lewis,S. and Rubin,G.M.
TITLE      BDGP/HMI Drosophila EST Project
JOURNAL    Unpublished (1997)
COMMENT    G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 26 row: C column: 2
            High quality sequence stop: 463.
            Location/Qualifiers
FEATURES   source
            1..629
            /organism="Drosophila melanogaster"
            /db_xref="taxon:7227"
            /clone="SD02626"
            /clone_lib="SD Drosophila melanogaster Schneider L2 cell
            culture pOT2"
            /lab_host="DH5-alpha"
            /note="Vector: pOT2; Site_1: EcoRI; Site_2: XhoI; Sized
            fractionated cDNAs were directly ligated into pOT2.
            Plasmid cDNA library."
BASE COUNT 188 a 132 c 133 g 176 t
ORIGIN

alignment_scores
            Quality: 171.50 Length: 219
            Ratio: 1.466 Gaps: 9
            Percent Similarity: 53.425 Percent Identity: 30.137

alignment_block:
US-09-301-906-15 x AI531540 ..
Align seg 1/1 to: AI531540 from: 1 to: 629
1955 AsnSerThrAlaThrIleThrAlaAsnValGlySerSerGluAspIleAs 1971
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
31 AATCCGTCCTCACTGCTACCTCGCGAGCGGTAGAGGAAAGACGACTTCAT 80
1971 nMetAlaValLysLysArgAspProAsnLeuGluGlyLeuAsnSerAlaT 1988
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
81 TGATCGTGTCACATAAATTTACCCAAA...GAAGCAATCGCGATTGCG 127
1988 hrThrValAsnSerArg...ValValAsnPhetIleValArgGlyMetTyr 2003
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
128 AAAAAATATACTCGCACCTCTGTGCAAGTTATCTGCTCCACCCGAAAG 177
2004 LysArg.....ValLeuValAspGluValHisMetMetHisG1 2016
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
178 AACCGAAAAACCGATAGTATCTTTATTGACGAAGCCCTTGATGTCCTATCC 227
2016 nGlyLeuLeuGlnLeuGlyValPheAlaThrGlyAlaSerGluGlyLeuP 2033
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
228 AGAGCGTGTGTTTTATGCTGCTGTGATGCTGTGTCGCTCAAAAGTTAGAT 277
2033 hePheGlyAspIleAsnGlnIleProPheIleAsnArgGluLysValPhe 2049
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
278 TTCCTCGTGACACTTGCAAAATTCATTTGTAACCGTTCTCCGAGTTC 327

```

```

2050 ArgMetAsp.....CysAlaValPheValProLysLysGluSerVa 2063
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
328 AAGATGAGTTACGATCGTTTGGAAAAAATGTCCTCAATTTGTCGGTTCACT 377
2063 lValTyrThrSerLysSerTyrArgCysProLeuAspValCysTyrLeuL 2080
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
378 TTTTATC.....TCGTACCGCTGCTCTGTCGATGTCGTTTAT.... 414
2080 euSerSerMetThrValArgGlyThrGluLysCysTyrProGluLysVal 2096
||| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
415 .....AGGTTTACCGACGACTACCGTGAAGTGAACAA 447
2097 ValSerGlyLysAspLysProValValArgSerLeuSerLysArgPro... 2112
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
448 AAGACGAGAAAGAT.....GTAAGTTTGATGAGTGCGCCGACGG 488
2113 .....IleGlyThrThrAspAspValAlaG 2121
489 TGTTAATACCTGTAGTTTGTCCGAATATCGACATTGATGAGTCCCGC 538
2121 luIleAsnAlaAspValTyrLeuCysMetThrGlnLeuGluLysSerAsp 2137
::: |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
539 TCCGTAAGGAAGTTAAATATCTCGTCTTTACAAAGCGAAAAAACCGAA 588
2138 MetLysArgSerLeuLysGlyLysGlyLysGluThrProValMetThrVa 2154
::: |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
589 CTTGCCATG.....AAGAAATATATATGTCTCGACAGT 620
2154 lHisGlu 2156
621 ACATGAA 627
seq_name: gb_est21:AI532043
seq_documentation_block: 636 bp mRNA EST 18-MAR-1999
LOCUS AI532043
DEFINITION SD03401.5prime SD Drosophila melanogaster Schneider L2 cell culture
            pOT2 Drosophila melanogaster cDNA clone SD03401 5prime, mRNA
            sequence.
ACCESSION  AI532043
VERSION     AI532043.1 GI:4446178
KEYWORDS   EST.
SOURCE     fruit fly.
ORGANISM   Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
            1 (bases 1 to 636)
            Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C., Brokstein
            ,P., Lewis,S. and Rubin,G.M.
            BDGP/HMI Drosophila EST Project
            Unpublished (1997)
            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 34 row: A column: 1
            High quality sequence stop: 505.
            Location/Qualifiers
FEATURES   source
            1..636
            /organism="Drosophila melanogaster"
            /db_xref="taxon:7227"
            /clone="SD03401"
            /clone_lib="SD Drosophila melanogaster Schneider L2 cell
            culture pOT2"
            /lab_host="DH5-alpha"
            /note="Vector: pOT2; Site_1: EcoRI; Site_2: XhoI; Sized
            fractionated cDNAs were directly ligated into pOT2.
            Plasmid cDNA library."
BASE COUNT 189 a 145 c 127 g 175 t
ORIGIN

```

alignment_scores:
Quality: 162.00 Length: 151
Ratio: 1.761 Gaps: 6
Percent Similarity: 60.927 Percent Identity: 33.775

alignment_block:
US-09-301-906-15 x AI532043 ..

Align seg 1/1 to: AI532043 from: 1 to: 636

1936 AlaProProGlyClyClyLysThrThrLeuValAspGluPheVally 1952
191 GCTGACCGCGTACTGGAANAACGACATGGATCCCTGAAGAACACGTCGC 240
1952 s.....SerProAsnSerThrAlaThrIleThrAlaAsnValGlySerS 1967
241 TTATTTTTCGAGACAATCCGTCCTCACTGTATATCTCGCGACGGTAGAGGAA 290
1967 erGluAspIleAsnMetAlaValLysLysArgAspProAsnLeuGluGly 1983
291 AAGACGACTTCATTGATCGTGTCAATAAAATTTACCCAAA...GAAGGC 337
1984 LeuAsnSerAlaThrThrValAsnSerArg...ValValAsnPheIleVa 1999
338 AATCGCGATTTCGCAAAAATATACTATCGCACTCTTGCAAGTTATCTGCT 387
1999 lArgGlyMetTyrlYsArg.....ValLeuValAspGluValH 2012
388 CCACCCCCAAAAGAACGAAACCGATGATCTTTATGACGAGCGCT 437
2012 IsMetMethIscInGlyLeuLeuInLeuGlyValPheAlaThrGlyAla 2028
438 TGATGTCCCATCCAGGAGCTGTTTTTATGCTGTGTTGGATGCTCGTSCG 487
2029 ScrGluGlyLeuPhePheClyAspIleAsnGlnIleProPheIleAsnAr 2045
488 TCANAAGTTAGATTCTCGGTGACAACTTGCAATTCATTTGTAACCG 537
2045 gGluLysValPheArgMetAsp.....CysAlaValPheValProL 2059
538 TTCTCCCGCAGTTCAGATGAGTTAGATCGTTTGCAAAAATGGTCCCAA 587
2059 ysLysGluSerValValTyThrSerLysSerTyArgCysProLeuAsp 2075
588 TTGTCGGTTCACCTTTTATC.....TCGTACCGCTGCTCTCGAT 628
2076 Val 2076
629 GTC 631

seq_name: gb_est21:AI531508

seq_documentation_block:
LOCUS AI531508 482 bp mRNA EST 18-MAR-1999
DEFINITION SD02578.5prime SD Drosophila melanogaster Schneider L2 cell culture
SD02 Drosophila melanogaster cDNA clone SD02578 5prime, mRNA
sequence.
ACCESSION AI531508
VERSION AI531508.1 GI:4445643
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 482)
AUTHORS Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C., Brokstein
P., Lewis, S. and Rubin, G.M.
TITLE BDGP/HIMI Drosophila EST Project
JOURNAL Unpublished (1997)
COMMENT Contact: Harvey, D.

G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
Plate: 25 row: G column: 6
High quality sequence stop: 447.
Location/Qualifiers
1. 482
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="SD02578"
/clone_lib="SD Drosophila melanogaster Schneider L2 cell
culture SD02"
/lab_host="DHS-alpha"
/note="vector: pOT2; Site:1: EcoRI; Site:2: XhoI; Sized
fractionated cDNAs were directly ligated into pOT2.
Plasmid cDNA library."
BASE COUNT 138 a 100 c 106 g 138 t
ORIGIN

alignment_scores:
Quality: 159.00 Length: 174
Ratio: 1.767 Gaps: 7
Percent Similarity: 51.724 Percent Identity: 32.184

alignment_block:
US-09-301-906-15 x AI531508 ..

Align seg 1/1 to: AI531508 from: 1 to: 482

2023 ValPheAlaThrGlyAlaSerGluGlyLeuPhePheGlyAspIleAsnG 2039
17 GTTGGATGCTGCTGGTGGTCAAAAGTTAGATTTCTCGGTGACAACTTGA 66
2039 nIleProPheIleAsnArgGluLysValPheArgMetAsp.....C 2053
67 AATTCATTGTTAAACCGTTCTCCGCGATTCAAGATGAGTTACGATCGTT 116
2053 ysAlaValPheValProLysLysGluSerValValTyThrSerLysSer 2069
117 TCGAAAAATGGTCCCAATTGTCGGTTCACCTTTTATC.....TCG 157
2070 TyrArgCysProLeuAspValCysTyrlLeuLeuSerSerMetThrValAr 2086
158 TACCGCTGCTGCTGCTGCTGCTTTAT.....AG 186
2086 gGlyThrGluLysCysTyrlProGluLysValValSerGlyLysAspLysP 2103
187 GTTACCGACGACTACCGTGAAGTGAACAAAAAAGCAGGAAAGAT.... 232
2103 roValValArgSerLeuSerLysArgPro..... 2112
233GTAGTTGATGAGTGCCTGCGACGGTGTAAATACCTGTAGTTT 277
2113IleGlyThrThrAspAspValAlaGluIleAsnAlaAspValTy 2127
278 GTCCGNAATATGACATTGAATGAGGTCCCGCTCCGTAAGGAAGTTAAATA 327
2127 rLeuCysMetThrGlnLeuGluLysSerAspMetLysArgSerLeuLysG 2144
328 TCTCGTCTTTACACAGCGCAAAACCGAACTTGCCATG..... 367
2144 lYlGlyLysGluThrProValMetThrValHisGluAlaGlnGlyLys 2160
368AAGAAATATAATGTCTGCACAGTACATGAATTCCAAGGCAAA 409
2161 ThrPheSerAspValValLeuPheArgThrLysLys...AlaAspAspSe 2176
410 GAGGCTAAACCATATACCTCTGCAGACTTAACCATATCCACAGACGGA 459
2176 rLeuPheThrLysGlnProHis 2183


```
BASE COUNT      178 a 129 c 125 g 159 t
ORIGIN
    Plasmid cDNA library."
alignment_scores:
    Quality: 158.00      Length: 150
    Ratio: 1.951        Gaps: 6
    Percent Similarity: 54.000    Percent Identity: 33.333
alignment_block:
US-09-301-906-15 x AI530886 ..
Align seg 1/1 to: AI530886 from: 1 to: 591
2069 SerTyrArgCysProLeuValCysTyrLeuLeuSerSerMetThrVa 2085
|||||
66 TCGTACCGTTCCTGCTCATGCTGTTAT..... 95
2085 lArgGlyThrGluLysCysTyrProGluLysValValSerGlyLysAspL 2102
|||||
96 .AGGTTTACCGGACGACTACCGTGAAGTGAACAAAGGAGGAGGAGGAGGAT. 143
2102 ysProValValArgSerLeuSerLysArgPro..... 2112
|||||
144 .....GTAGGTTTGAGTTCGCGCGGCGGCGGTGTTAATACCTGTAGG 185
2113 .....lIleGlyThrThrAspAspValAlaGluIleAsnAlaAspVa 2126
|||||
186 TTTGTCGGAATATCGACATGTAATGAGTCCCGCTCCCGTAAGGAAGTTAA 235
2126 lTyrLeuCysMetThrGlnLeuGluLysSerAspMetLysArgSerLeuL 2143
|||||
236 ATATCTCGTCTTTACAAAGCGCGGAAACCGCAACTTGCCATG..... 278
2143 ysGlyLysGlyLysGluThrProValMetThrValHisGluAlaGlnGly 2159
|||||
279 .....AAGAAATATAATGTCTCGACAGTACATGAATTCGAAGGC 317
2160 LysThrPheSerAspValValLeuPheAlaThrLysLys...AlaAspAs 2175
|||||
318 AAAGAGGCTAAACCATATACCTCGTCAGACTTACCAATATCCAGACAG 367
2175 pSerLeuPheThrLysGlnProHisIleLeuValGlyLeuSerArgHisT 2192
|||||
368 GGAGCTATTCTCGGACATCCGTCAGCTCTAGTCTGCTTAACTCGTCATA 417
2192 hrArgSerLeuValThrAlaAlaLeuSerSerLysLeuAspLysVal 2208
|||||
418 CTGAAGAACTTGTCTAT.....TACACACGCGAAGACGGATGATCTGTG 461
seq_name: qb_est21:AI543590
seq_documentation_block:
LOCUS      AI543590      580 bp      mRNA      EST      22-MAR-1999
DEFINITION SD10387.5prime SD Drosophila melanogaster Schneider L2 cell culture
           p022 Drosophila melanogaster cDNA clone SD10387 5prime, mRNA
           sequence.
ACCESSION      AI543590
VERSION        AI543590.1      GI:4460963
KEYWORDS      EST.
SOURCE        fruit fly.
ORGANISM      Drosophila melanogaster
              Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
              Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
              Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE      1 (bases 1 to 580)
              Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C., Brokstein
              P., Lewis,S. and Rubin,G.M.
              BDGP/HIMI Drosophila EST project
              Unpublished (1997)
              Contact: Harvey, D.
              G. M. Rubin-Molecular and Cell Biology
```

```
University of California Berkeley
539 LSA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: 103 row: H column: 3
High quality sequence stop: 501.
Location/Qualifiers
    1..580
    /organism="Drosophila melanogaster"
    /db_xref="taxon:7227"
    /clone_lib="SD Drosophila melanogaster Schneider L2 cell
    culture p022"
    /lab_host="DH5-alpha"
    /note="Vector: p022; Site 1: EcoRI; Site 2: XhoI; Sized
    fractionated cDNAs were directly ligated into p022.
    Plasmid cDNA library."
BASE COUNT      170 a 122 c 128 g 160 t
ORIGIN
alignment_scores:
    Quality: 155.00      Length: 160
    Ratio: 1.598        Gaps: 7
    Percent Similarity: 60.625    Percent Identity: 33.125
alignment_block:
US-09-301-906-15 x AI543590 ..
Align seg 1/1 to: AI543590 from: 1 to: 580
1942 LysThrThrThrLeuValAspGluPheValLys.....SerProAsnSe 1956
|||||
3 AAACGACATGGATCTCTGAAGAACACGTCGCTTATTTGACACACATCC 52
1956 rThrAlaThrIleThrAlaAsnValGlySerSerGluAspIleAsnMeta 1973
|||||
53 GTCCACTGTCTACTCGCGACGGTAGGGGAAAAAGACGACTTCATTGATC 102
1973 laValLysLysArgAspProAsnLeuGluGlyLeuAsnSerAlaThrThr 1989
|||||
103 GTGTCAATAAATTTACCCAAA...GAAGGCAATCGGATTTGCAAAA 149
1990 ValAsnSerArg...ValValAsnPheIleValArgGlyMetTyrLysAr 2005
|||||
150 ATATACTATCGCAGCTCTTGCAGATTATCTGCCACCCGCAAGAACCG 199
2005 g.....ValLeuValAspGluValHisMetMetHisGlnGlyl, 2018
|||||
200 AAAACCGGATAGTATCTTTATTGACGAAGCCTTGATGTCCTCCAGGAG 249
2018 euLeuGlnLeuGlyValPheAlaThrGlyAlaSerGluGlyLeuPhePhe 2034
|||||
250 CAGTTTTTATCTGTTTGGATGCTGGTGGGTCGCAAAAGTTAGATTCTC 299
2035 GlyAspIleAsnGlnIleProPheIleAsnArgGluLysValPheArgMe 2051
|||||
300 GGTGACAACTTGCATAATTCATTTGTAACCGTTCTCCGAGTTCAAGAT 349
2051 tAsp.....CysAlaValPheValProLysLysGluSerValVal 2065
|||||
350 GAGTTACGATCGTTTGGAAAAAATGTCCTCCCAATTTGTCGGTTCAC 399
2065 yThrSerLysSerTyrArgCysProLeuAspValCysTyrLeuLeuSe 2081
|||||
400 TC.....TCGTACCCGCTGCTGTCGTGCTGTTATAGGTTTACC 440
2081 rSerMetThrValArgGlyThrGluLys 2090
|||||
441 GACGACTACCGTGAAGTGACAAAAAAG 468
seq_name: qb_est21:AI542293
```

```

seq_documentation_block:
LOCUS AI542293 573 bp mRNA EST 22-MAR-1999
DEFINITION SD08556.5prime SD Drosophila melanogaster Schneider L2 cell culture
pot2 Drosophila melanogaster cDNA clone SD08556 5prime, mRNA
sequence.
ACCESSION AI542293
VERSION AI542293.1 GI:4459566
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 573)
AUTHORS Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C., Brokstein
,P., Lewis,S. and Rubin,G.M.
TITLE BDGP/HIMI Drosophila EST Project
JOURNAL Unpublished (1997)
COMMENT G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: 85 row: E column: 8
High quality sequence stop: 472.
FEATURES
source
location/Qualifiers
1..573
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="SD08556"
/clone_lib="SD Drosophila melanogaster Schneider L2 cell
culture pot2"
/lab_host="DH5-alpha"
/note="Vector: pot2; site_1: EcoRI; site_2: XhoI; sized
fractionated cDNAs were directly ligated into pot2.
Plasmid cDNA library."
BASE COUNT 173 a 116 c 122 g 162 t
ORIGIN
alignment_scores:
Quality: 149.50 Length: 145
Ratio: 1.942 Gaps: 5
Percent Similarity: 53.103 Percent Identity: 33.103
alignment_block:
US-09-301-906-15 x AI542293 ..
Align seg 1/1 to: AI542293 from: 1 to: 573
2006 ValLeuValAspGluValHisMetMetHisGlnGlyLeuGlnLeuG1 2022
:::|||||::: ||| ||| :::
138 ATCTTTATTGACGAGCCTTGATGTCCTCCATCCAGGAGCTGTTTATGC 187
2022 yValPheAlaThrGlyAlaSerGluGlyLeuPhePheGlyAspIleAsnG 2039
|||||::: |||||||::: |||||||:::
188 TGTTCGATGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 237
2039 lntleProPheIleAsnArgGluLysValPheArgMetAsp..... 2052
|||||:::|||||::: |||:::|||||:::
238 AATTCATTTGTAACCGTTCCTCCGAGTTCAGATGAGTTCAGATCGT 287
2053 CysAlaValPheValProLysLysGluSerValValValThrSerLysSe 2069
:::||||| |||:::
288 TTCGAAAAATGTCCTCCATTCGTCGTCACATTTTATC.....TC 328
2069 rTyrArgCysProLeuAspValCysTyrLeuLeuSerSerMetThrValA 2086
|||||:::|||||::: |||
329 GTACCGCTGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 357
2086 rgGlyThrGluLysCysTyrProGluLysValValSerGlyLysAspLys 2102
|| |||:::

```

```

358 GGTTTACCGAGACTACCGTGAAGTGACAAAAAGCAGGAAAAAGAT... 404
2103 ProValValArgSerLeuSerLysArgPro..... 2112
|||:::|||| |||
405 .....GTAGTTTGATGAGTGCGCGCAGCGTGTAAATACCTGAGTT 448
2113 .....IleGlyThrThrAspAspValAlaGluIleAsnAlaAspValT 2127
|||||:::|||||::: |||:::
449 TGTCGGAATATCGACATGATGAGTCCCGCTCCGTAAGGAAGTTAAAT 498
2127 yrLeuCysMetThrGlnLeuGlyLysSerAspMet 2138
|||||:::||||| |||||||:::
499 ATCTCGTCCTTACACAGCGGAAAAACCGAAGCTT 533
seq_name: gb_est21:AI542833
seq_documentation_block:
LOCUS AI542833 448 bp mRNA EST 22-MAR-1999
DEFINITION SD09311.5prime SD Drosophila melanogaster Schneider L2 cell culture
pot2 Drosophila melanogaster cDNA clone SD09311 5prime, mRNA
sequence.
ACCESSION AI542833
VERSION AI542833.1 GI:4460206
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 448)
AUTHORS Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C., Brokstein
,P., Lewis,S. and Rubin,G.M.
TITLE BDGP/HIMI Drosophila EST Project
JOURNAL Unpublished (1997)
COMMENT G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: 93 row: A column: 11
High quality sequence stop: 340.
FEATURES
source
location/Qualifiers
1..448
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="SD09311"
/clone_lib="SD Drosophila melanogaster Schneider L2 cell
culture pot2"
/lab_host="DH5-alpha"
/note="Vector: pot2; site_1: EcoRI; site_2: XhoI; sized
fractionated cDNAs were directly ligated into pot2.
Plasmid cDNA library."
BASE COUNT 131 a 95 c 94 g 128 t
ORIGIN
alignment_scores:
Quality: 144.00 Length: 145
Ratio: 1.618 Gaps: 6
Percent Similarity: 61.379 Percent Identity: 33.103
alignment_block:
US-09-301-906-15 x AI542833 ..
Align seg 1/1 to: AI542833 from: 1 to: 448
1955 AsnSerThrAlaThrIleThrAlaAsnValGlySerSerGluAspIleAs 1971
|||:::||||| |||:::|||||:::
5 AATCCGTCCTGTCATCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 54
1971 nMetAlaValLysLysArgAspProAsnLeuGluGlyLeuAsnSerAlaT 1988
|||||:::||||| |||:::

```

```

55 TCATCGTGTCAATAAAATTTACCCAAA...GAAGCAATCCGATTTGC 101
1988 hrThrValAsnSerArg...ValValAsnPheIleValArgGlyMetTyr 2003
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
102 AAAAAATATACTACGCACTCTTGCAGTTATCTGCTCCACCCCGAAGAAG 151
2004 lysArg.....ValLeuValAspGluValHisMetMetHisGI 2016
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
152 AACCGAANAACGATAGTATCTTTATTGACGAAGCCTGATGCCATCC 201
2016 nclyLeuLeuGlnLeuGlyValPheAlaThrGlyAlaSerGluGlyLeuP 2033
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
202 AGGAGCTGTTTTTATGCTGTGTGTGATGCTGTCGTCGCAAAAGTTAGAT 251
2033 hePheGlyAspIleAsnGlnIleProPheIleAsnArgGluLysValPhe 2049
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
252 TTCTCGGTGACAACTTGCAATTCATTGTGAACCGTTCTCCGAGTTC 301
2050 ArgMetAsp.....CysAlaValPheValProLysLysGluSerVa 2063
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
302 AAGATGAGTTACGATCGTTTGGAAAAAATGGTCCCAATTGTGGTTCACT 351
2063 lValTyrThrSerLysSerTyrArgCysProLeuAspValCysTyrLeu. 2079
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
352 TTTTATC.....TCGTACCGCTGTCCTGTCGATGCTGTTTATAGGT 392
2080 LeuSerSerMetThrValArgGlyThrGluLys 2090
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
393 TTACCGACGACTACCGTGAAGTGACAAAAAAG 425
seq_name: gb_est40:AW160301

seq_documentation_block: 571 bp mRNA EST 08-NOV-1999
LOCUS AW160301
DEFINITION EST290159 L. pennellii trichome, Cornell University Lycopersicon
pennellii cDNA clone cLPT1011, mRNA sequence.
ACCESSION AW160301
VERSION AW160301.1 GI:6279835
KEYWORDS EST.
SOURCE Lycopersicon pennellii.
ORGANISM Lycopersicon pennellii.
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids
I; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 571)
AUTHORS Alcala,J., Vrebalov,J., White,R., Matern,A.L., Lakey,J., Holt,I.E.,
Liang,F., Hansen,T.S., Upton,J., Ronning,C.M., Craven,M.B., Fujii
,C.Y., Bowman,C.L., Nierman,W., Fraser,C.M., Venter,J.C., Martin
,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from wild tomato (Lycopersicon pennellii)
trichomes
JOURNAL Unpublished (1999)
COMMENT Contact: David Frisch
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@CLEMSON.EDU
5 prime sequence.
FEATURES
    source
        Location/Qualifiers
            1..571
                /organism="Lycopersicon pennellii"
                /db_xref="taxon:28526"
                /clone="cLPT1011"
                /clone_lib="L. pennellii trichome, Cornell University"
                /tissue_type="trichome"
                /dev_stage="mixed stages"
                /lab_host="SOLR"
                /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; Leaves of various stages were shaken in liquid
nitrogen, shearing off trichomes. This procedure yielded a
mixture of cells highly enriched for trichomes, with minor

```

```

BASE COUNT 160 a 110 c 152 g 148 t 1 others
ORIGIN
alignment_scores:
    Quality: 139.50      Length: 159
    Ratio: 1.516        Gaps: 4
    Percent Similarity: 57.862      Percent Identity: 29.560
alignment_block:
US-09-301-906-15 x AW160301 ..
Align seg 1/1 to: AW160301 from: 1 to: 571
1937 ProProGlyGlyGlyLysThrThrLeuValAspGluPheValLysSe 1953
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
52 CCACCGGTAGCTCTGCTAAAGTTGTACTGGTAGACGG...GTTCCGGG 98
1953 rProAsnSerThrAlaThrIleThrAlaAsnValGlySerSerGluAspI 1970
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
99 ATGCGGGAAGACGAAGGAAATTTCTATCGAAGGTAAACTTTGATGAGACC 148
1970 leAsnMetAla.....ValLysLysArgAspPro 1979
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
149 TGATTCTCGTACCCGCTCGTCAGGCTGCTGAGATGATCAGAGAAGACGCG 198
1980 AsnLeuGluGlyLeuAsnSerAlaThrThrValAsnSerArgValVal.. 1995
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
199 AACTCGTCGGGAGTAATAGTACCCACAAAGGATAATGTCAGAACCGTCGA 248
1996 AsnPheIleVal.....ArgGlyMetTyrLysArgV 2006
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
249 TTCATTATCATGAAATACGGGAAGGAGTGCCTGCCAGTTCARAAAGCG 298
2006 alLeuValAspGluValHisMetMetHisGlnGlyLeuLeuGlnLeuGly 2022
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
299 TTTTATTGATGAAGGATTGATGCTGCACACTGCTGTGTGAATTTCTTG 348
2023 ValPheAlaThrGlyAlaSerGluGlyLeuPhePheGlyAspIleAsnGI 2039
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
349 GTAAACAATCTCTCTGCGCATATTCATATGATATACGGAGACACCCACA 398
2039 nileProPheIleAsnArgGluLysValPheArgMetAspCysAlaValP 2056
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
399 AATTCATATATACACAGATACCGGATTTCCGTACCTGCTCATTTT 448
2056 heValProLysLysGluSerValValTyrThrSerLysSerTyrArgCys 2072
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
449 CAAAGCTGGAGGTGGACGAGGTGGAGACACCGCAAACTACGTTGCGTGT 498
2073 ProLeuAspValCysTyrLeuLeuSer 2081
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
499 CCGGCCGATGTCACCTATTTCTTAAAT 525
seq_name: gb_est21:AF531122
seq_documentation_block: 607 bp mRNA EST 18-MAR-1999
LOCUS AF531122
DEFINITION SD02020.5prime SD Drosophila melanogaster Schneider L2 cell culture
pot2 Drosophila melanogaster cDNA clone SD02020 5prime, mRNA
sequence.
ACCESSION AF531122
VERSION AF531122.1 GI:4445246
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 607)
AUTHORS Harvey,D., Hong,L., Evans,Holm,M., Pendleton,J., Su,C., Brokstein
,P., Lewis,S. and Rubin,G.M.

```

```

TITLE      BDGP/HIMI Drosophila EST Project
JOURNAL    Unpublished (1997)
COMMENT    Contact: Harvey, D.
           G. M. Rubin-Molecular and Cell Biology
           University of California Berkeley
           539 LSA, Berkeley, CA 94720-3200, USA
           Fax: 510 643 9947
           Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
           Plate: 20 row: B column: 8
           High quality sequence stop: 526.
FEATURES   Location/Qualifiers
            1..607
            /organism="Drosophila melanogaster"
            /db_xref="taxon:7227"
            /clone="SD02020"
            /clone_lib="SD Drosophila melanogaster Schneider L2 cell
            culture pot2"
            /lab_host="DH5-alpha"
            /note="Vector: pot2; Site.1: EcoRI; Site.2: XhoI; Sized
            fractionated cDNAs were directly ligated into pot2.
            Plasmid cDNA library."
BASE COUNT 185 a 137 c 119 g 166 t
ORIGIN

alignment_scores:
    Quality: 139.00      Length: 123
    Ratio: 1.782         Gaps: 4
    Percent Similarity: 63.415      Percent Identity: 33.333

alignment_block:
US-09-301-906-15 x AI531122 ..
Align seg 1/1 to: AI531122 from: 1 to: 607

1936 AlapropGlyGlyGlyLysThrThrLeuValAspGluPheVally 1952
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
183 GCTGGACCGGTACTGGAAAGACATGATGCTCTGAAGAACACGTGCC 232
::: ||| ::::::::::||| ||||| ||||| |||||
1952 s.....SerProAsnSerThrAlaThrIleThrAlaAsnValGlySers 1967
::: ||| ::::::::::||| ||||| ||||| |||||
233 TTATTGTGCAGCAATCCGCCACTGTCATCTACTCGCGACGTAGAGGAA 282
::: ||| ::::::::::||| ||||| ||||| |||||
1967 erGluAspIleAsnMetAlaValLysLysArgAspProAsnLeuGluGly 1983
::: ||||| ::::::::::||| ||||| ||||| |||||
283 AAGACGACTTCATGTGCTGCTCAATAAATTTACCCAAA...GAAGGC 329
::: ||| ::::::::::||| ||||| ||||| |||||
1984 LeuAsnSerAlaThrThrValAsnSerArg...ValValAsnPheIleVa 1999
::: ||| ::::::::::||| ||||| ||||| |||||
330 ATCTGGGATTGTGAAAATAATATCTATCGCACTCTTGCAAGTTATCTGCT 379
::: ||| ::::::::::||| ||||| ||||| |||||
1999 lArgGlyMetTyrLysArg.....ValLeuValAspGluValH 2012
::: ||| ::::::::::||| ||||| ||||| |||||
380 CCACCCGGAAGAACCGAAACCGAATAGATATCTTTATTGACGAGCCT 429
::: ||| ::::::::::||| ||||| ||||| |||||
2012 isMetMetHisGlnGlyLeuLeuGlnLeuGlyValPheAlaThrGlyAla 2028
||| ||| ||| ::::::::::||| ||||| ||||| |||||
430 TGAATGCCCATCCAGGAGCTGTTTATTTATGCTGCTGTTGGATGCTGCTG 479
|||::: |||::: |||::: |||::: |||::: |||::: |||:::
2029 SerGluGlyLeuPhePheGlyAspIleAsnGlnIleProPheIleAsnAr 2045
|||::: |||::: |||::: |||::: |||::: |||::: |||:::
480 TCAAAAGTTAGATTCTCGGTGACAACTTGCAAAATTCATTGTGAACCG 529
|||::: |||::: |||::: |||::: |||::: |||::: |||:::
2045 gGluLysValPheArgMet 2051
|||::: |||::: |||::: |||::: |||::: |||::: |||:::
530 TTCCTCCGAGTTCAGATG 548

seq_name: gb_est21:AI542359

seq_documentation_block:
LOCUS      AI542359          346 bp      mRNA      22-MAR-1999
DEFINITION SD08652.5prime SD Drosophila melanogaster Schneider L2 cell culture
            pot2 Drosophila melanogaster cDNA clone SD08652 5prime, mRNA

```

```

sequence.
ACCESSION  AI542359
VERSION    AI542359.1  GI:4459732
KEYWORDS   EST.
SOURCE     fruit fly.
ORGANISM   Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE  1 (bases 1 to 346)
AUTHORS    Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C., Brokstein
            P., Lewis,S. and Rubin,G.M.
TITLE      BDGP/HIMI Drosophila EST Project
COMMENT    Unpublished (1997)
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 86 row: E column: 4
            High quality sequence stop: 292.
FEATURES   Location/Qualifiers
            1..346
            /organism="Drosophila melanogaster"
            /db_xref="taxon:7227"
            /clone="SD08652"
            /clone_lib="SD Drosophila melanogaster Schneider L2 cell
            culture pot2"
            /lab_host="DH5-alpha"
            /note="Vector: pot2; Site.1: EcoRI; Site.2: XhoI; Sized
            fractionated cDNAs were directly ligated into pot2.
            Plasmid cDNA library."
BASE COUNT 92 a 66 c 82 g 106 t
ORIGIN

alignment_scores:
    Quality: 135.00      Length: 89
    Ratio: 2.411         Gaps: 3
    Percent Similarity: 62.921      Percent Identity: 40.449

alignment_block:
US-09-301-906-15 x AI542359 ..
Align seg 1/1 to: AI542359 from: 1 to: 346

2006 ValLeuValAspGluValHisMetMetHisGlnGlyLeuLeuGlnLeuG1 2022
::: ||| ||| ||| ||| ||| ||| ||| ||| |||
26 ATCTTTATTGACGAAGCCTTGATGTCCTCCATCCAGGAGCTGTTTTTATGC 75

2022 yValPheAlaThrGlyAlaSerGluGlyLeuPhePheGlyAspIleAsnG 2039
|||::: |||::: |||::: |||::: |||::: |||::: |||:::
76 TGTGTGATGCTGCTGGTCGTCAAAAGTTAGATTTCTCGGTGACAACTTGC 125

2039 lnIleProPheIleAsnArgGluLysValPheArgMetAsp..... 2052
|||::: |||::: |||::: |||::: |||::: |||::: |||:::
126 AAATTCCATTGTAAACCGTTCTCCGAGTTCAAGATGAGTTACGATCGT 175

2053 CysAlaValPheValProLysLysGluSerValValThrThrSerLysSe 2069
|||::: |||::: |||::: |||::: |||::: |||::: |||:::
176 TTGAAAAAATGGTCCCAATGTCGGTTCTACCTTTTATC.....TC 216

2069 rTyrArgCysProLeuAspValCysTyrLeu.LeuSerSerMetThrVal 2085
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
217 GTACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 266

2086 ArgGlyThrGluLys 2090
::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
267 AAGTGAACAAAAAAG 281

seq_name: gb_est21:AI532485

```

```

seq_documentation_block: 488 bp mRNA EST 18-MAR-1999
LOCUS AI532485 SD04021.5prime SD Drosophila melanogaster Schneider L2 cell culture
DEFINITION SD04021.5prime SD Drosophila melanogaster cDNA clone SD04021 5prime, mRNA
sequence.
ACCESSION AI532485
VERSION AI532485.1 GI:4446620
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 488)
AUTHORS Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C., Brokstein
P., Lewis, S. and Rubin, G.M.
TITLE BDGP/HMI Drosophila EST Project
JOURNAL Unpublished (1997)
COMMENT Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: 40 row: B column: 9
High quality sequence stop: 366.
FEATURES
source
1..488
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="SD04021"
/lab_host="SD Drosophila melanogaster Schneider L2 cell
culture pOT2"
/note="Vector: pOT2; Site_1: EcoRI; Site_2: XhoI; Sized
fractionated cDNAs were directly ligated into pOT2.
Plasmid cDNA library."
BASE COUNT 156 a 107 c 102 g 123 t
ORIGIN
alignment_scores:
Quality: 128.50 Length: 148
Ratio: 1.586 Gaps: 7
Percent Similarity: 54.730 Percent Identity: 31.081
alignment_block:
US-09-301-906-15 x AI532485 ..
Align seg 1/1 to: AI532485 from: 1 to: 488
2098 SerGlyLysAspLysProValValArgSerLeuSerLysArgPro..... 2112
:::|||||:::
35 GCAGGAAGAAT.....GTAGTTTGATGAGTCGCGGACGGTGT 75
2113 .....IleGlyThrThrAspValAlaGluI 2122
|||||:::
76 TAATACCTGTAGTTTGCAGATATCGACATTGAATGAGGTCGCGTCC 125
2122 leAsnAlaAspValThrLeuGlyMetThrGlnLeuGluLysSerAspMet 2138
:::|||||:::
126 GTAAGGAAGTTAAATATCTCGCTTTTACACAGCGCAAAACCGAAGCTT 175
2139 LysArgSerLeuLysGlyLysGlyLysGluThrProValMetThrValHi 2155
|||||:::
176 GCCATG.....AAGAAATATATGTCGACAGATACA 207
2155 sGluAlaGlnGlyThrPheSerAspValValLeuPheArgThrLysL 2172
|||||:::
208 TCAATTCCAAGCAAGAGGCTAAACCATATACCTCGTCAGACTTAACC 257
2172 ys....AlaAspAspSerLeuPheThrLysGlnProHisIleLeuValGly 2187
:::|||||:::

```

```

258 AATATCCACAGACGGACGATATTCTTGGGACATCCGTACGTCTTAGTCGCT 307
2188 LeuSerArgHisThrArgSerLeuValTyrAlaAlaLeuSerSerLysLe 2204
|||||:::
308 TTAACCTCGTACATCGAGAACTTGTCTAT.....TACACACGCAGAAC 351
2204 uAspAspLysVal.....GlyThrTyrIleSerA 2214
|||||:::
352 GGATGATCTGTGTGTAGAAAAATCAAAAGTGACACGTTTATGTCTTTC 401
2214 sp....AlaSerProGlnSerValSerAspAlaLeuLeuHisThr 2227
:::|||||:::
402 AACTCGCTCACTGACAAATACGACGCTCTCTTAAAAACTACA 445
seq_name: gb_est21:AI531921
seq_documentation_block: 429 bp mRNA EST 18-MAR-1999
LOCUS AI531921 SD03228.5prime SD Drosophila melanogaster Schneider L2 cell culture
DEFINITION SD03228.5prime SD Drosophila melanogaster cDNA clone SD03228 5prime, mRNA
sequence.
ACCESSION AI531921
VERSION AI531921.1 GI:4446056
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 429)
AUTHORS Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C., Brokstein
P., Lewis, S. and Rubin, G.M.
TITLE BDGP/HMI Drosophila EST Project
JOURNAL Unpublished (1997)
COMMENT Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: 32 row: C column: 4
High quality sequence stop: 371.
FEATURES
source
1..429
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="SD03228"
/lab_host="SD Drosophila melanogaster Schneider L2 cell
culture pOT2"
/note="Vector: pOT2; Site_1: EcoRI; Site_2: XhoI; Sized
fractionated cDNAs were directly ligated into pOT2.
Plasmid cDNA library."
BASE COUNT 139 a 90 c 92 g 108 t
ORIGIN
alignment_scores:
Quality: 126.00 Length: 145
Ratio: 1.658 Gaps: 6
Percent Similarity: 52.414 Percent Identity: 31.034
alignment_block:
US-09-301-906-15 x AI531921 ..
Align seg 1/1 to: AI531921 from: 1 to: 429
2074 LeuAspValCysTyrLeuLeuSerSerMetThrValArgGlyThrGlu 2090
:::|||||:::
9 GTCGATGTCGTTTAT.....AGGTTTACCGACA 37
2090 sCysTyrProGluLysValValSerGlyLysLysProValValArg 2107
:::|||||:::

```

```
38 CTACCGTGAAGTGAACAAAAAGCAGGAAAGAT.....GTAGGTT 78
2107 erLeuSerLysArgPro.....IleGly 2114
      ::::: |||
79 TGATGAGTGC GCGACGGTGTAAATACCTGTAGGTTTGCCGAATATCG 128
2115 ThrThrAspValAlaGluIleAsnAlaAspValTyrLeuCysMetTh 2131
      ||| ::::: ||| ::|
129 ACATTGAATGAGTCCCGCTCCGTAAGGAAGTTAAATATCTCGTCTTTAC 178
2131 rGlnLeuGluLysSerAspMetLysArgSerLeuLysGlyLysGlyLys 2148
      ||| ||||| :::::
179 ACAAGCGGAAAAAACCGAAGCTTCCCATG.....AAGA 210
2148 lUthrProValMetThrValHisGluAlaGlnGlyLysThrPheSerAsp 2164
      :: ||| ||||| |||||
211 AATATAATGTCTCGACAGTACATGAATTCCAAGGCAAGAGGCTAAACCC 260
2165 ValValLeuPheArgThrLysLys...AlaAspSerLeuPheThrLy 2180
      ::| ||| ||| ::::: ::|
261 ATATACCTCTCGACGCTTAACCAATATCCACAGACGACGATATCTTGCG 310
2180 sGlnProHisTleLeuValGlyLeuSerArgHisThrArgSerLeuValT 2197
      ::::: ||||| ::||| ::||| ::||| ::|||
311 ACATCCGTAGTCTTAGTCGCTTTAACTCGTCATCTGAGAACTTGCT 360
2197 yrAlaAlaLeuSerSerLysLeuAspLysVal 2208
      || ::| ::| ||||| |||
361 AT.....TACACACGACGAGAACGATGATCTTG 389
```

This Page Blank (just to)

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/cgnl_7/prodata1/lna/6.COMB.seq:US-08-770-544-1 +		5247.00	10066.20	0.0	4173 !
/cgnl_7/prodata1/lna/6.COMB.seq:US-09-081-320-2 +		290.00	485.29	1.8e-19	6485 !
/cgnl_7/prodata1/lna/6.COMB.seq:US-09-081-320-1 +		290.00	480.65	3.2e-19	8743 !
/cgnl_7/prodata1/lna/5B.COMB.seq:US-08-687-559-2 +		278.00	462.33	3.4e-18	6395 !
/cgnl_7/prodata1/lna/6.COMB.seq:US-09-259-741-3 +		278.00	462.26	3.4e-18	6425 !
/cgnl_7/prodata1/lna/6.COMB.seq:US-09-337-751-3 +		278.00	462.26	3.4e-18	6425 !
/cgnl_7/prodata1/lna/6.COMB.seq:US-09-259-741-5 +		278.00	462.21	3.4e-18	6446 !
/cgnl_7/prodata1/lna/6.COMB.seq:US-09-037-751-5 +		278.00	462.21	3.4e-18	6446 !
/cgnl_7/prodata1/lna/6.COMB.seq:US-09-259-741-4 +		278.00	462.14	3.4e-18	6475 !
/cgnl_7/prodata1/lna/6.COMB.seq:US-09-037-751-4 +		277.00	462.14	3.4e-18	6475 !
/cgnl_7/prodata1/lna/6.COMB.seq:US-09-259-741-1 +		277.00	460.40	4.3e-18	6395 !
/cgnl_7/prodata1/lna/6.COMB.seq:US-09-037-751-1 +		277.00	460.40	4.3e-18	6395 !
/cgnl_7/prodata1/lna/6.COMB.seq:US-09-259-741-2 +		277.00	460.29	4.4e-18	6439 !
/cgnl_7/prodata1/lna/6.COMB.seq:US-09-037-751-2 +		277.00	460.29	4.4e-18	6439 !
/cgnl_7/prodata1/lna/6.COMB.seq:US-08-840-316-4 +		187.50	285.76	2.3e-08	7168 !
/cgnl_7/prodata1/lna/PCTUS.COMB.seq:PCT-US-0935-08849A-4 +		187.50	285.76	2.3e-08	7168 !
/cgnl_7/prodata1/lna/PCTUS.COMB.seq:PCT-US-0935-08849-4 +		187.50	285.76	2.3e-08	7168 !
/cgnl_7/prodata1/lna/6.COMB.seq:US-08-478-507-10 +		171.00	253.89	1.4e-06	7171 !
/cgnl_7/prodata1/lna/6.COMB.seq:US-08-478-507-6 +		169.50	250.94	2.0e-06	7195 !
/cgnl_7/prodata1/lna/backfiles1.seq:5183745-5 +		165.00	249.03	2.6e-06	4649 !
/cgnl_7/prodata1/lna/backfiles1.seq:5183745-5 +		165.00	243.96	4.9e-06	6443 !
/cgnl_7/prodata1/lna/6.COMB.seq:US-09-090-793-13 +		163.50	224.08	6.3e-05	19227 !
/cgnl_7/prodata1/lna/6.COMB.seq:US-09-090-793-12 +		163.50	212.65	0.0003	40138 !
/cgnl_7/prodata1/lna/5B.COMB.seq:US-08-680-517-13 +		150.50	224.72	5.8e-05	3666 !
/cgnl_7/prodata1/lna/5B.COMB.seq:US-08-682-517-14 +		150.50	224.72	5.8e-05	3666 !
/cgnl_7/prodata1/lna/5B.COMB.seq:US-08-682-517-7 +		150.50	222.62	7.6e-05	4197 !
/cgnl_7/prodata1/lna/5B.COMB.seq:US-08-682-517-8 +		150.50	222.62	7.6e-05	4197 !
/cgnl_7/prodata1/lna/5B.COMB.seq:US-08-853-659A-6 +		140.00	190.54	0.0046	8967 !
/cgnl_7/prodata1/lna/5B.COMB.seq:US-08-853-659A-9 -		140.00	190.54	0.0046	8967 !
/cgnl_7/prodata1/lna/5B.COMB.seq:US-08-853-659A-64 +		140.00	190.54	0.0046	8967 !
/cgnl_7/prodata1/lna/5B.COMB.seq:US-08-853-659A-67 -		140.00	190.54	0.0046	8967 !
/cgnl_7/prodata1/lna/5B.COMB.seq:US-07-920-281C-1 +		140.00	186.66	0.0076	11517 !
/cgnl_7/prodata1/lna/5B.COMB.seq:US-08-853-659A-2 +		140.00	174.80	0.0348	24701 !
/cgnl_7/prodata1/lna/5B.COMB.seq:US-08-853-659A-3 -		140.00	174.80	0.0348	24701 !
/cgnl_7/prodata1/lna/5B.COMB.seq:US-08-853-659A-60 +		140.00	174.80	0.0348	24701 !
/cgnl_7/prodata1/lna/5B.COMB.seq:US-08-853-659A-61 -		140.00	174.80	0.0348	24701 !
/cgnl_7/prodata1/lna/5A.COMB.seq:US-08-409-995-3 +		138.50	194.58	0.0028	5738 !
/cgnl_7/prodata1/lna/6.COMB.seq:US-08-685-467-3 +		138.50	194.58	0.0028	5738 !
/cgnl_7/prodata1/lna/5A.COMB.seq:US-08-711-119A-1 +		138.00	160.98	0.2048	46899 !
/cgnl_7/prodata1/lna/5B.COMB.seq:US-08-843-530B-3 +		137.50	194.25	0.0029	5175 !

101 TCCTGGTTAGCATACCACTTCGGGCATCTTCGACAACTTGAACCTTCGG 150
 898 GlyAlaPheAspIleSerLysGluThrPheGlyArgLysLeuLysAsnse 914
 151 GCGCGTTTCGAGGTTCTTAAAGAAATCTCCAGGAGTTACGTTTCGAG 200
 914 rArgLeuArgValPheSerArgAlaIleValGluAspSerIleLysValM 931
 201 TCGTTTGGCGGTATTCTAGGGTATTGTGGAGGATACGATCAAGGTTA 250
 931 eLysAlaMetLysThrGluAspGlyLysProLeuProIleThrGluAsp 947
 251 TCAAGGGCATGAATCAGAGATGTAAACCACTCCCTATAGCCGAGGAT 300
 948 SerValTyrAlaPheIleMetGlyAsnValSerAsnValHisCysThrAr 964
 301 TCCGTGTACCGCTTCATGACAGCAATATGTCAACGTTTCATTGCACTAG 350
 964 gAlaGlyLeuLeuGlyGlySerLysAlaThrValValSerSerValSerL 981
 351 GGCTGGTTTGTTCGGGGGCTCAAAAGGCTTGGCGGCTTCTTTAGCTGTGA 400
 981 ysGlyLeuValAlaArgGlyAlaAlaThrLysAlaPheSerGlyIleThr 997
 401 AGGTTGCAGCTTCACCGCTACTGGAACAAACCTTTTTCAGGTCACAA 450
 998 SerPhePheSerThrGlySerLeuPheTyrAspArgGlyLeuThrGluAs 1014
 451 TCCTTTCTTCGGCGGTGGTCTGTTTACGATGAAGCTTGACGCCCGG 500
 1014 pGluArgLeuAspAlaLeuValArgThrGluAsnAlaIleAsnSerProV 1031
 501 AGAGAGGCTTGTACACTAACCGCGCGTGAACATGCTGTGAATTCACCTG 550
 1031 alGlyIleLeuGluThrSerArgValAlaValSerLysValValAlaGly 1047
 551 TAGGCTCTTGAAGACCTCGAGCTTCGGTGGCGAAGCGGTGTTTCCGGA 600
 1048 ThrLysGluPheTrpSerGluValSerLeuAsnAspPheThrThrPheVa 1064
 601 ACGAAAGCTTTCTGTCAAGATTGTTCATTGGAGGACTTCACCACTTCGT 650
 1064 lLeuArgAsnLysValLeuIleGlyIlePheValAlaSerLeuGlyAlaA 1081
 651 CATAAAAATAGGTTGCTTATGTGTTTACTCTTTTCCATGGCTCTCA 700
 1081 laProIleAlaTrpLysTyrArgArgGlyIleAlaAlaAsnAlaArgArg 1097
 701 CTCGGTGTCTGGAGGTACAGAGGAATATCGCGCAACTGGCGTGGAT 750
 1098 TyrAlaGlySerSerTyrGluThrLeuSerSerLeuSerSerGlnAlaAl 1114
 751 GTTTTCCACCGTGTCTGGGTACCGCGCCATCGGTTTACAATGTC 800
 1114 aGlyGlyLeuArgGlyLeuThrSerSerThrValSerGlySerLeuV 1131
 801 TAGTGGAGGAAGGTCGTAGTGTGGTACGCTGCTCGTGGCGGCTTAACAG 850
 1131 alValArgArgPheSerSerAlaValThrValThrArgAlaThrVal 1147
 851 TGACTCGAGGAGGCTATCTTCGGCGGTGGGTGACCAAGAAATACAGT 900
 1148 AlaLysArgGlnValProLeuAlaLeuLeuSerPheSerThrSerTyrAl 1164
 901 GCTAGGCGTCAGGTACCATTTGGCGTGTGCTTTTCGTTTCCACGCTCTACGC 950
 1164 aIleSerGlyCysSerMetLeuGlyIleTrpAlaHisAlaLeuProArgH 1181
 951 AGTCAGTGGTGCACCTTGTAGGTATTGTGGGCTCATGCTCTCCCTAGGC 1000
 1181 lsLeuMetPhePhePheGlyLeuGlyThrLeuLeuGlyAlaArgAlaSer 1197
 1001 ATTTGATGTTCTCTTTGGCCCTAGGACCGCTCTTCGGGGTGTAGTCCAGT 1050

1198 AlaAsnThrTrpLysPheGlyGlyPheSerAsnAsnTrpCysAlaValPr 1214
 1051 ACCAATTCCTGGTCTGGGGCTATACGAACAGTCTGTTCCACGTACC 1100
 1214 oGluValValTrpArgGlyLysSerValSerSerLeuLeuLeuProIle 1231
 1101 GGAATTAACCTTGGAGGAGGAGTTACAGATCTTTATGTCGCCCAAGCAG 1150
 1231 hrLeuGlyValSerLeuIleIleArgGlyLeuLeuAsnAspThrIlePro 1247
 1151 CTTTAGGTATTTCTCTGTTGTCGCGGGTGTAAAGTGAAGTGTGCCA 1200
 1248 GlnLeuAlaTyrValProValGluGlyArgAsnValTyrAspGluTh 1264
 1201 CAACTAACGTACGTACCGCGGATTAAGTGCGAATGTTTATGATCAGCG 1250
 1264 rLeuArgTyrTyrArgAspPheAspTyrAspGluGlyAlaGlyProSerG 1281
 1251 ACTAAATTTTATCCGACTTTTACTATGACATGATGTCAGGCCCATCCG 1300
 1281 lyThrGlnHisGluAlaValProGlyAspAspAsnAspGlySerThrSer 1297
 1301 GGACGGCTGCTCAAGCGATCCTGGA...ACCAATACTTCGGATACTTCT 1347
 1298 SerValSerSerTyrAspValValThrAsnValArgAspValIglyIleSe 1314
 1348 TCGGTTTCTCT...GACGATGGTTGCT 1373
 1314 rThrAsnGly...GluValThrGlyGluG 1323
 1374 CGCTAGTCGGGTGGCTTCGAGCGCGCTTGAGCGAGTCCACCCATG 1423
 1323 LuGluThrHisSerProArg...SerValGlnTyrThrTyrValGluGlu 1338
 1424 CTGTTGATGAATCACCAGGGTGTGTTGAGTTCTGCTACAGAGAAGCT 1473
 1339 ...GluValAlaProSerAlaAlaValAlaGluArgGlnGlyAsp... 1352
 1474 GTAGATGAACATCCGCGCTGTGCTGAAGCTGAAGTTGAAGGATCTAAT 1523
 1353 ...ProSerGlySerGlyThrAlaAspAla...M 1362
 1524 AACACCACTTGTACAGCTGCTTGAAGTCCGCCCGCTAGTCTCTGAAG 1573
 1362 eAlaPheValGluSerValLysLysGlyValAspValPheHisGln 1378
 1574 CTGGAGCGCGCCCAACGTCGAGGAGGTTTCCCGAGGTT... 1614
 1379 GlnSerSerGlyGluThrAlaArgGluValGluValAspGlyLysGlyLe 1395
 1615 ...GAAGCTGAGAAATGTTCCGAGGTCATCGTTGAC... 1647
 1395 uLeuProGluSerValValGlyGluAlaProThrGlnGluArgGlyArgA 1412
 1648 ...GTCTAGTTCA...GAACCGCGGTACAAGAGTCTCTTGAAT 1687
 1412 laAlaAspGlyAsnThrAlaGlnThrAlaValAsnGluGlyAspArgGlu 1428
 1688 CAACCAATGGT...GTCCAAGCTGCAAGACTGAA...GAG 1722
 1429 ProValGlnSerSerLeuValSerSerProGlnAlaAspIleProLysVa 1445
 1723 GTTGTGACGGCGCACATGTGGAGCT...GGSGT 1754
 1445 lThrGlnSerGluValHisAlaGlnLysGluValLysGlnGluValPro. 1461
 1755 AGCTAAATCAGAAGTG...AGTCAACGTGTGTTCTCGCAAGTACCAG 1801
 1462 ...LeuAlaThrValSerGlyAla... 1468
 1802 CACATGAAGCTGGTCTTGAGGCATCTAGTGGCGCGCTGCTGGAGCCATG 1851

1469ThrProIleValAspGluLysProAlaProSerVal...Th 1481
1852 CAAGTTTCTGTGCAGTAGCCGTAGAGAAACTGTTTATCTCTCGAGAA 1901
1481 rThrArgGlyValLysIleIleAspLysGlyLysAlaValAlaHisVala 1498
1902 GCGCGTGACCTAAAGCGGTAGATAGGCGCAAGCGCGTCTGCACGCAA 1951
1498 laGluLysLysGlnValGlnValGlnProLysGlnArgSerLeuThr 1514
1952 AGGAAGTCAGAAGATGACCGGTAAAGACGTTACACAGCGGGCTCTAAAA 2001
1515 IleAsnGluGlyLysAlaGlyLysGlnLeuCysMetPheArgThrCysSe 1531
2002 ATTAGTGAGGATACCGTCTCAGGAATGTGTGATGTTTGAACGGTTC 2051
1531 rCysGlyValGlnLeuAspValTyrAsnGluAlaThrIleAlaThrArgP 1548
2052 CTGCGCGTGCAGTTGGACGTGTACAATGAAGCACCATCCGCACATAGT 2101
1548 heSerAsnAlaPheThrPheValIleAspAsnLeuLysGlyArgSerAlaVal 1564
2102 TCTCAATGCGTTTACCTTTGTCGATAGCTTGAAGGGAGGAGTGGCGTC 2151
1565 PhePheSerLysLeuGlyGlyGlyTyrThrTyrAsnGlyGlySerHisVa 1581
2152 TTTTCTCAAGCTGGGTAGGGGTATACCTATAATGGTGGTACCCATGT 2201
1581 lSerSerGlyTrpProArgAlaLeuGluAspIleLeuThrAlaIleLysT 1598
2202 TTTCATCAGGGTGCCCTCGTCCCTAGAGGATATCTTAACGGCAATTAAGT 2251
1598 yrProSerValPheAspHisCysLeuValGlnLysTyrLysMetGlyGly 1614
2252 ACCCAAGCGCTTCGACCACTGTTTAGTGCAGAGTACAAGATGGGTGA 2301
1615 GlyValProPheHisAlaAspAspGluGluCysTyrProSerAspAsnPr 1631
2302 GCGGTACCATTCACCGCTGATGACGAGGAGTGTATCCATCAGATTAACC 2351
1631 oIleLeuThrValAsnLeuValGlyLysAlaAsnPheSerThrLysCysA 1648
2352 TATCTTGACGGTCAATCTCGTGGGAAGGCAAACTCTCGACCTAAGTGCA 2401
1648 rGlyGlyGlyLysValMetValIleAsnValAlaSerGlyAspTyrPhe 1664
2402 GGAAGGGTGGTAAGGTCATGGTCATAAACGCTAGCTTCGGGTGACTATTT 2451
1665 LeuMetProCysGlyPheGlnArgThrHisLeuHisSerValAsnSerII 1681
2452 CTTATGCCCTTGGGTGTTTCAAGGAGCGCACTTGCATTACAGTAACCTCAT 2501
1681 eAspGluGlyArgIleSerLeuThrPheArgAlaThrArgValPheG 1698
2502 CGACGAAGGGCGCATCAGTTTGACGTTACGGGCAACTCGCGCGTCTTTG 2551
1698 lyValGlyArgMetLeuGlnLeuAlaGlyGlyValSerAspGluLysSer 1714
2552 GTCTAGCGAGGATGTTGCAGTTAGCGCGCGCGTGTCCGATGAGAAGTCA 2601
1715 ProGlyValProAsnGlnProGlnSerGlnGlyAlaThrArgThrII 1731
2602 CCAAGGTGTTCCAAACACGCAACACAGAGCCAAAGGTGCTACCAAGAACAT 2651
1731 eThrProLysSerGlyGlyLysAlaLeuSerGluGlySerGlyArgGluV 1748
2652 CACACCAAAATCGGGGGCAAGGCTCTATCTGAGGGAAGTGTAGGGAAG 2701
1748 alCysGlyArgSerThrTyrSerIleTrpCysGluGlnAspTyrValArg 1764
2702 TCNAAGGGAGGTGCATACCTACTCGATATGGTGCACAAGATTACGTTAGG 2751
1765 LysCysGluTrpLeuArgAlaAspAsnProValMetAlaLeuGluProAs 1781

2752 AAGTGTAGTGGCTCAGCGCTGATAATCCAGTGATGGCTCTTAAACCTGG 2801
1781 pTyrThrProMetThrPheGluValValLysThrGlyThrSerGluAspA 1798
2802 CTACACCCCAATGACATTTGAAGTGGTTAAAGCCGGACCTCTCAAGATG 2851
1798 laValValGluTyrLeuLysTyrLeuAlaIleGlyIleGluArgThrTyr 1814
2852 CCGTCGTGGAGTACTTGAAGTATCTGGCTATAGGCATTTGGAGGACATAC 2901
1815 ArgAlaLeuLeuMetAlaArgAsnIleAlaValThrAlaGluGlyVa 1831
2902 AGGCGGTGCTTATGGCTAGAAATATTGCCGTACTACCGCCGAAGGTGT 2951
1831 lLeuLysValProAsnGlnValTyrGluSerLeuProGlyPheHisValT 1848
2952 TCTCAAAAGTACCTAAATCAAGTTTATGATCACTACCGGGCTTTTCACGTTT 3001
1848 yrLysSerGlyThrAspLeuIlePheHisSerThrGlnAspGlyLeuArg 1864
3002 ACAAGTGGGCACAGATCTCATTTTTCATTCAACACAAGACGGCTTGGCT 3051
1865 ValArgAspLeuProTyrValLeuIleAlaGluLysGlyIlePheThrLy 1881
3052 GTGAGAGACCTACCGTACGTATTCATAGCTGAGAAAGGTATTTTATCAA 3101
1881 sGlyLysAspValAspAlaValValAlaLeuGlyAspAsnLeuPheValC 1898
3102 GGCACAAAGATGTCAGCGGGTAGTAGCTTGGCGCAAACTCTGTCGGTAT 3151
1898 ysAspAspIleLeuValPheHisAspAlaIleAsnLeuIleGlyAlaLeu 1914
3152 GTGATGATATATTGGTTTTCATGATGCTATTAATTGATGGGTGCACGT 3201
1915 LysValAlaArgCysGlyMetValGlyGluSerPheLysSerPheGluTy 1931
3202 AAAGTTCCTCGATGTTGTTGGTGGTGAATCATTTAAGTCGTTTCAATA 3251
1931 rLysCysTyrAsnAlaProProGlyGlyGlyLysThrThrThrLeuVala 1948
3252 CAATGCTATTAATGCTCCCGCAGGTGGCGGTAGACGACGATGCTAGTGG 3301
1948 spGluPheValLysSerProAsnSerThrAlaThrIleThrAlaAsnVal 1964
3302 ACGAAATTTGTCAAGTCACCAATAGCAGCGCCACCATTAACGGCTAACGTG 3351
1965 GlySerSerGluAspIleAsnMetAlaValLysLysArgaspProAsnLe 1981
3352 GGAAGTTCTCGAGGACATAAATATGGCGGTGAAGAAGAGATCCGAATTT 3401
1981 uGluGlyLeuAsnSerAlaThrThrValAsnSerArgValValAsnPheI 1998
3402 GGAAGGTCTCAACAGTCTACCACAGTTAACTCCAGGGTGGTTAACTTTTA 3451
1998 leValArgGlyMetTyrLysArgValLeuValAspGluValHisMetMet 2014
3452 TTGTACAGGGGAATGTAATAAAGGGTTTTTGGTGGATGAGGTGTACATGATG 3501
2015 HisGlnGlyLeuLeuGlnLeuGlyValPheAlaThrGlyAlaSerGluG 2031
3502 CATCAAGGCTTACTACAACATAGCGCTCTTCGCAACCGCGCGCTCGGAAG 3551
2031 yLeuPhePheGlyAspIleAsnGlnIleProPheIleAsnArgGluLysV 2048
3552 CCTCTTTTTTGGAGACATAAATCAGATACCATTCATTAACCGGGGAGG 3601
2048 alPheArgMetAspCysAlaValPheValProLysLysGluSerValVal 2064
3602 TGTTTAGGATGGATTGTGCTGTTTCTTCCAAAGAGGAAGGCGTTGTA 3651
2065 TyrThrSerLysSerTyrArgCysProLeuAspValCysTyrLeuLeuSe 2081

```
3652 TACACTTCTAAATCATACAGGTGTCGGTTAGATGTTTGCCTACTGTTGTC 3701
2081 rSerMetThrValArgGlyThrGluLysCysTyrProGluLysValValS 2098
|||||
3702 CTTCAATGACCGTAAGGGGAACGAAAGGTTTACCTGAAAGGTCGTTA 3751
2098 erGlyLysAspLysProValValArgSerLeuSerLysArgProIleGly 2114
|||||
3752 GCGGTGAAGACAACACAGTAGTAAGATCCTGCTCCAAAAGGCAATTGGA 3801
2115 ThrThrAspValAlaGluIleAsnAlaAspValTyrLeuCysMetTh 2131
|||||
3802 ACCATGATCAGCTAGCTGAATAAACCGTGACCTGACTTGTGCATGAC 3851
2131 rGlnLeuGluLysSerAspMetLysArgSerLeuLysGlyLysG 2148
|||||
3852 CCAGTTGGAGAAGTCGGATATCAAGAGGTCGTTGAAGGGAAGAAAG 3901
2148 luThrProValMetThrValHisGluAlaGlnGlyLysThrPheSerAsp 2164
|||||
3902 AAACACCAAGTATGACAGTGCATGAAGCACAGGGAACAAATTCAGTGAT 3951
2165 ValValLeuPheArgThrLysLysAlaAspSerLeuPheThrLysG 2181
|||||
3952 GTGGTATGTTTAGCAGGAAGAACCGGATGACTCCCTATTCACTAAACA 4001
2181 nProIleLeuValGlyLeuSerArgHisThrArgSerLeuValTyrA 2198
|||||
4002 ACCGCATATACTTGTGTTGTTGTCGAGACACACACGCTCACCTGGTTATG 4051
2198 laAlaLeuSerSerLysLeuAspAspLysValGlyThrTyrIleSerAsp 2214
|||||
4052 CCGCTCTGAGCTCAGAGTTGGAGATAAGTGGCGCACATATATTAGCGAC 4101
2215 AlaSerProGlnSerValSerAspAlaLeuLeuHisThrPheAlaProAl 2231
4102 GCGTCGCCCTCAATCAGTATCCGACGCTTTCCTTCACACGCTTCGCCCGGC 4151
2231 aGlyCysPheArgGlyIle 2237
|||||
4152 TGGTTGCTTCGAGGTATA 4170
```

seq_name: /cgnl_7/ptodata/1/ina/6_COMB.seq:US-09-081-320-2

```
seq_documentation_block:
; Sequence 2, Application US/09081320
; Patent No. 6093544
; GENERAL INFORMATION:
; APPLICANT: Consalves, Dennis
; APPLICANT: Meng, Baozhong
; TITLE OF INVENTION: RUPESTRIS STEM PITTING ASSOCIATED VIRUS
; TITLE OF INVENTION: NUCLEIC ACIDS, PROTEINS, AND THEIR USES
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/081,320
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/047,147
; FILING DATE: 20-MAY-1997
; PRIOR APPLICATION DATA:
```

```
; APPLICATION NUMBER: US 60/069,902
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/1722
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6485 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-09-081-320-2
```

alignment_scores:

Quality:	290.00	Length:	142
Ratio:	2.736	Gaps:	3
Percent Similarity:	74.648	Percent Identity:	41.549

alignment_block:

US-09-301-906-15 x US-09-081-320-2 ..

Align seg 1/1 to: US-09-081-320-2 from: 1 to: 6485

```
1554 PheValAspAsnLeuLysGlyArgSerAlaValPhePheSerLysLeuG 1570
|||||
2197 TTTAAAGACAGGCTGAGAGCGGGTGTCATCTTCTCAAAACACAGG 2246
1570 y...GluGlyTyrThrTyrAsnGlySerHisValSerSerGlyTrp 1586
|||||
2247 AATTTCATGTCATAGTTACAATGGTGGGACCAACAAGCTTAGGTTGGC 2296
1586 roArgAlaLeuGluAspIleLeuThrAlaIleLysTyrProSerValPhe 1602
|||||
2297 CAAAGTTTCATGGATCAGATTCCTAAGCTCCACCTGGTGGAGCTAATTACTAC 2346
1603 AspHisCysLeuValGlnLysTyrLysMetGlyGlyValProPheH 1619
::: |||||
2347 AATTTCATGCTGGCTCAGATCTATGAGGAAATTTCAAAATTGGCTCTTCA 2396
1619 saLaAspAspGluGluCysTyrProSerAspAsnProIleLeuThrValA 1636
|||||
2397 TAAGGATGATGAGAGTTGCTATGAAATTGGGCACAAAGTTTTCACCTGTTA 2446
1636 snLeuValGlyLysAlaAsnPheSer...ThrLysCysArgLys..... 1649
|||||
2447 ATTTAATCGGCTCAGCAACTTTCACCTATTAGTAAGTCGCGAAATTTGGTT 2496
1650 GlyGlyLysValMetValIleAsnValAlaSerGlyAspTyrPheLeuMe 1666
|||||
2497 GGGGGTAAATCATTTGCAGCCTGACAAATTTGGCCAAATAGTTTTTCGAAAT 2546
1666 tProCysGlyPheGlnArgThrHisLeuHisSerValAsnSerIleAspG 1683
|||||
2547 GCCTAGGGGCATGCAATTACTTCATGGGGTTTCCCAATTGTACGC 2596
1683 luClyArgIleSerLeuThrPheArg 1691
|||||
2597 CAGGGCGGTATCGCTGACCTTAGG 2622
```

seq_name: /cgnl_7/ptodata/1/ina/6_COMB.seq:US-09-081-320-1

```
seq_documentation_block:
; Sequence 1, Application US/09081320
; Patent No. 6093544
; GENERAL INFORMATION:
; APPLICANT: Consalves, Dennis
; APPLICANT: Meng, Baozhong
```

```

; TITLE OF INVENTION: RUPESTRIS STEM PITTING ASSOCIATED VIRUS
; TITLE OF INVENTION: NUCLEIC ACIDS, PROTEINS, AND THEIR USES
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESSES:
; ADDRESSSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/081,320
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/047,147
; FILING DATE: 20-MAY-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/069,902
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/1722
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8743 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-09-081-320-1

alignment_scores:
  Quality: 290.00      Length: 142
  Ratio: 2.736        Gaps: 3
  Percent Similarity: 74.648      Percent Identity: 41.549

alignment_block:
US-09-301-906-15 x US-09-081-320-1  ..
Align seg 1/1 to: US-09-081-320-1 from: 1 to: 8743

1554 PheValAspAsnLeuLysGlyArgSerAlaValPheSerLysLeuG1 1570
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2258 TTTAAGACAGCGTGAGAGCGGGGTGCATCTTCTCAAAACCCAGG 2307
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1570 Y...GluGlyTyrThrTyrAsnGlyGlySerHisValSerSerTrpP 1586
| :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
2308 AATTTCATGTTCATAGTTACAAATGGTGGGAGCCACAAAGCTTAGG 2357
|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1586 roArgAlaLeuGluAspIleLeuThrAlaIleLysTyrProSerValPhe 1602
|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
2358 CAAAGTTTCAGGATCATGATCTTAAGCTCCACCTGGTGACGTAATTA 2407
|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1603 AspHisCysLeuValGlnLysTyrLysMetGlyGlyValProPheHi 1619
:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
2408 AATTGATGCTGGCTCAGATCTATGAGAAATTCAAAATGGCTCTTCA 2457
| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1619 sAlaAspAspGluCysTyrProSerAspAsnProIleLeuThrVala 1636
| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2458 TAAGGATGATGACAGTTGCTATGAAATGGGCACAAAGTTTGCAC 2507
|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1636 snLeuValGlyLysAlaAsnPheSer...ThrLysCysArgLys..... 1649

```

```

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2508 ATTTAATCGGCTCAGACACTTTCACATTATTAGTCCGAAATTTGGTT 2557
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1650 GlyGlyLysValMetValIleAsnValAlaSerGlyAspTyrPheLeuMe 1666
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2558 GGGGTAAATCATTCAGCCTGACCAATTTGGGCCAAATGAGTTTTCGAAAT 2607
1666 tProCysGlyPheGlnArgThrHisLeuHisSerValAsnSerIleAspG 1683
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2608 GCCTAGGGGCGATGCAATGCAATTTACTTCATGGGTTTCCAATTTACGC 2657
1683 luGlyArgIleSerLeuThrPheArg 1691
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2658 CAGGGCGGATGCTGCTGACCTTTAGG 2683

seq_name: /cgnl_7/ptodata/1/ina/58_COMB.seq:US-08-687-559-2

seq documentation_block:
; Sequence 2, Application US/08687559
; Patent No. 5955647
; GENERAL INFORMATION:
; APPLICANT: Fitch, John H.
; APPLICANT: Beachy, Roger N.
; TITLE OF INVENTION: METHOD FOR USING TOBACCO MOSAIC VIRUS TO
; TITLE OF INVENTION: OVERPRODUCE PEPTIDES AND PROTEINS
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESSES:
; ADDRESSSEE: Fish & Richardson, P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,559
; FILING DATE: No. 5955647ember 18, 1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/01467
; FILING DATE: 03-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Learn, June M.
; REGISTRATION NUMBER: 31,238
; REFERENCE/DOCKET NUMBER: 07302/011001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-678-5070
; TELEFAX: 619-678-5099
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6395 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; IMMEDIATE SOURCE:
; CLONE: TWV
; US-08-687-559-2

alignment_scores:
  Quality: 278.00      Length: 1895
  Ratio: 0.401        Gaps: 60
  Percent Similarity: 36.570      Percent Identity: 16.570

alignment_block:
US-09-301-906-15 x US-08-687-559-2  ..
Align seg 1/1 to: US-08-687-559-2 from: 1 to: 6395

```


1344 ATTCGATCGAGGCTA.....ATCATTAACGGTGTGAC 1375
972 sAlaThrValValSerSerValSerLysGlyLeuValAlaAArgGlyAlaA 989
1376 AGCGAGTCCGAATGGGATGTGACAAATCTTTGTACAA..... 1415
989 laThrLysAlaPheSerGlyIleThrSerPhePheSerThrGlySerLeu 1005
1416TCCTTGTCATGACG 1430
1006 PheTyrAspArgGlyLeuThrGluAspGluArgLeuAspAlaLeuValAr 1022
1431 TTTTAC..... 1436
1022 gThrGluAsnAlaIleAsnSerProValGlyIleLeuGluThrSerArgV 1039
1437CTGCATACTAAGCTTCCGTTCTTAAGGATGAC...T 1470
1039 aAlaValSerLysValAlaGlyThrLys.....GluPhe 1051
1471 TACTGATTAGCAAGTTTAGTCTCGGTTGAAACCGGTGCCAGCATGTG 1520
1052 TrpSerGluValSerLeuAsnAspPheThrThrPheValLeuArgAsnLy 1068
1521 TGGGATGAGATTCGCTG..... 1538
1068 sValLeuIleGlyIlePheValAlaSerLeuGlyAlaAlaProIleAlaT 1085
1538 1538
1085 rPlpYstYrArgArgGlyIleAlaAlaAsnAlaArgArgTyrAlaGlySer 1101
1538 1538
1102 SerTyrGluThrLeuSerSerLeuSerSerGlnAlaGlyGlyLeuAr 1118
1538 1538
1118 gGlyLeuThrSerSerThrValSerGlyGlySerLeuValAlaArgArg 1135
1538 1538
1135 lYPheSerSerAlaValThrValThrArgAlaThrValAlaLysArgGln 1151
1538 1538
1152 ValProLeuAlaLeuLeuSerPheSerThrSerTyrAlaIleSerGlyCy 1168
1538 1538
1168 sSerMetLeuGlyIleTrpAlaHisAlaLeuProArgHisLeuMetPheP 1185
1538 1538
1185 hepGlyLeuGlyThrLeuLeuGlyAlaArgAlaSerAlaAsnThrTrp 1201
1538 1538
1202 LysPheGlyGlyPheSerAsnAsnTrpCysAlaValProGluValValTr 1218
1539GCGTTTGGGACCGATTCCCTCCGTGAAGAGGCGCTT 1579
1218 pArgGlyLysSerValSerSerLeuLeuLeuProIleThrLeuGlyValS 1235
1580 GAACAGGAACATTATCAGA.....GTGGCAGCGCAGC 1611
1235 erLeuIlelleArgGlyLeuLeuAsnAspThrIleProGlnLeuAlaTyr 1251
1612 CATTAGATGATCAG.....GTGCCTGATCTA...TAT 1640
1252 ValProProValGluGlyArgAsnValTyrAspGluThrLeuArgTyrTy 1268
|||

1641 GTG..... 1643
1268 rArgAspPheAspTyrAspGluGlyAlaGlyProSerGlyThrGlnHisG 1285
1644ACCTTCCACG 1653
1285 luAlaValProGlyAspAsnAspGlySerThrSerSerValSerSer 1301
1654 ACAGATTAA.....GTGACTGAG 1670
1302 TyrAspValValThrAsnValArgAspValIcIlyIleSerThrAsnGlyG 1318
1671 TACAAGCCCTCTGTGGACATGCGCTTGACATTAGGAGAGATGGA 1720
1318 uValThrGlyGluGluGluThrHisSerProArgSerValGlnTyrThrT 1335
1721 AGAAACG.....GAAGTATGTACAATG 1743
1335 yrValGluGluGluValAlaProSerAlaAlaValAlaGluArgGlnGly 1351
1744 CACTTTTCAGAG..... 1754
1352 AspProSerGlySerGlyThrAlaAspAlaMetAlaPheValGluSerVa 1368
1755TTATCGGTGTTAAGGGAGTC 1774
1368 lLysLysGlyValAspAspValPheHisGlnGlnSerSerGlyGluThra 1385
1775 TGCAAAATTCGATGTTGATGTTTTTCCCATG.....T 1809
1385 laArgGluValGluValAspGlyLysGlyLeuLeuProGluSerValVal 1401
1810 GCCAATCTTTGGAAGTTGAC.....CCAATGACGGCAGCAAGGTTATA 1853
1402 GlyGluAlaProThrGlnGluArgGly.....ArgAl 1412
1854 GTCGCGGTGTCATGAGCAATGAGCGGTCTGACTCTCACATTTGNACGCC 1903
1412 aAlaAspGlyAsnThrAlaGlnThrAlaValAsnGluGlyAspArgGluP 1429
1904 TACTGAGCGCAATGTTGCG...CTAGCTTTACAGGATCAAGAGAGGCTT 1950
1429 roValGlnSerSerLeuValSerSerProGlnAlaAspIleProLysVal 1445
1951 CAGAAAGTGTCTTGGTAGTTACCTCAAGAGAAGTTGAAGAACCGCTCCATG 2000
1446 ThrGlnSerGluValHisAlaGlnLysGluValLysGlnGluValProle 1462
2001 AAGGGTTCG.....ATGCCAGAGGAGAGTTACATTT 2032
1462 uAlaThrValSerGlyAlaThrProIleValAspGluLysProAlaProS 1479
2033 AGCTGGTCTTGTGAGATCATCCG..... 2057
1479 erValThrThrArgGlyValLysIleIleAspLysGlyLysAlaValAla 1495
2057 2057
1496 HisValAlaGluLysLysGlnValGlnValGlnProLysGlnArgSe 1512
2058 GAGTCGCTCTATTCTAAGAAGGAGGATAGAGTCTTTAGAGCAGATTCA 2107
1512 rLeuThrIleAsnGluGlyLysAlaGlyLysGlnLeuCysMetPheArgT 1529
2108 TATGGCAACGGCAGATTCGTTAATTCGTAAAGCAGATGAGCTCGATT... 2153
1529 hrCysSerCysGlyValGlnLeuAspValTyrAsnGluAlaThrIleAla 1545
2154GTGTACACGGGTCCGCTTAAAGTT 2177
1546 ThrArgPheSerAsnAlaPheThrPheValAspAsnLeuLysGlyArgSe 1562
2178 CAGCAATGAAAAAC.....TTTATCGATAGCTGTGATGATCACT 2218

1562 rAlaValPhePheSerLysLeuGlyGluGlyTyrThrTyrAsnGlyGlys 1579
 : : : : :
 2219 ATCTGCT 2225
 1579 erHisValSerSerGlyTrpProArgAlaLeuGluAspIleLeuThrAla 1595
 : : : : :
 2226 2249
 1596 IleLysTyrProSerValPheAspHisCysLeuValGlnLysTyrLysMe 1612
 : : : : :
 2250 CTCAAAGATACAGCTGCTATTGAC 2293
 1612 tGlyGlyGlyValProPheHisAlaAspGluGluCysTyrProSera 1629
 : : : : :
 2294 TGCA 2297
 1629 sPAsnProIleLeuThrValAsnLeuValGlyLysAlaAsnPheSerThr 1645
 2297 2297
 1646 LysCysArgLysGlyGlyLysValMetValIleAsnValAlaSerGlyAs 1662
 : : : : :
 2298 2320
 1662 pTyrPheLeuMetProCysGlyPheGlnArgThrHisLeuHisSerVala 1679
 : : : : :
 2321 GTGGTTAATCAACCA 2336
 1679 snSerIleAspGluGlyArgIleSerLeuThrPheArgAlaThrArgArg 1695
 2336 2336
 1696 ValPheGlyValGlyArgMetLeuGlnLeuAlaGlyGlyValSerAspG1 1712
 2336 2336
 1712 uLysSerProGlyValProAsnGlnGlnProGlnSerGlnGlyAlaThrA 1729
 2336 2336
 1729 rgThrIleThrProLysSerGlyGlyLysAlaLeuSerGluGlySerGly 1745
 : : : : :
 2337 2378
 1746 ArgGluValLysGlyArgSerThrTyrSerIle 1758
 : : : : :
 2379 AGGAG 2410
 1758 sGluGlnAspTyrValArgLysCysGluTrpLeuArgAlaAspAsnProv 1775
 : : : : :
 2411 TCAGCAGGGTGTGGTGACATCGCATGATTGGAGAAGA 2447
 1775 alMetAlaLeuGluProAspTyrThrProMetThrPheGluValLys 1791
 2447 2447
 1792 ThrGlyThrSerGluAspAlaValValGluTyrLeuLysTyrLeuAlaI1 1808
 : : : : :
 2448 GTAGCTGTGACGCTGAGTCTGTGTTT 2485
 1808 eGlyIleGluArgThrTyrArgAlaLeuLeuMetAlaArgAsnIleAlaV 1825
 : : : : :
 2486 GCGAAACTCAGAACTCGCAGACTGCTTCGAAACCGAGAACCGCATG 2535
 1825 alThrThrAlaGluGlyValLeuLysValProAsnGlnValTyrGluSer 1841
 : : : : :
 2536 TCAGTAGCGCAAGGTTGTCTT 2558
 1842 LeuProGlyPheHisValTyrLysSerGlyThrAspLeuIlePheHis 1858
 2558 2558

1858 rThrGlnAspGlyLeuArgValArgAspLeuProTyrValLeuIleAlaG 1875
 2558 2558
 1875 LuLysGlyIlePheThrLysGlyLysAspValAspAlaValValAlaLeu 1891
 : : : : :
 2559 2570
 1892 GlyAspAsnLeuPheValCysAspIleLeuValPheHisAspAlaI1 1908
 2570 2570
 1908 eAsnLeuIleGlyAlaLeuLysValAlaArgCysGlyMetValGlyGlu 1925
 2570 2570
 1925 erPheLysSerPheGluTyrLysCysTyrAsnAlaProProGlyGlyGly 1941
 : : : : :
 2571 2582
 1942 LysThrThrThrLeuValAspGluPheValLysSerProAsnSerThrAl 1958
 : : : : :
 2583 AAA 2590
 1958 aThrIleThrAlaAsnValGlySerSerGluAspIleAsnMetAlaValL 1975
 : : : : :
 2591 AGAAATTCCTTCCAGGTTAATTTTGATCAAGATCTAATTTTAGTACCTC 2640
 1975 ysLys 1984
 : : : : :
 2641 GGAGCAAGCCCGGAAATGATCAGAAGACGTCGGAATTCCTCAGGGATT 2690
 1985 AsnSerAlaThr 1996
 : : : : :
 2691 ATTGTGCCCAAGCAAGACACACGTTAAACCGTTGATTCTTTTCATGATGA 2740
 1996 nPhe 2011
 : : : : :
 2741 TTTTGGAAAGCACACGCTGTCAAGTTCAGAGGTTATTTCATTGATGAAG 2790
 2011 alHisMetMetHisGlnGlyLeuLeuGlnLeuGlyValPheAlaThrGly 2027
 : : : : :
 2791 GGTTCATGTTGCATACTGTTGTGTTAAATTTCTTGTGGCGATGTCATTG 2840
 2028 AlaSerGluGlyLeuPhePheGlyAspIleAsnGlnIleProPheIleAs 2044
 : : : : :
 2841 TCGAAATTCATATGTTTACGAGACACACACAGATTCATATACATCAA 2890
 2044 nArgGluLysValPheArgMetAspCysAlaValPheValProLysLysG 2061
 : : : : :
 2891 TAGAGTTTCAGGATTCCCGTACCCGCCCATTTTGCCAAATTTGGAAGTTG 2940
 2061 LuSerValValTyrThrSerLysSerTyrArgCysProLeuAspValCys 2077
 : : : : :
 2941 ACGAGGTGGAGACACGACAACTACTCTCCGTTGTCAGCCGATGTCTACA 2990
 2078 TyrLeuLeuSerSerMetThrValArgGlyThrGluLysCysTyrProG1 2094
 : : : : :
 2991 CATTATCTGAAC 3016
 2094 uLysValValSerGlyLysAspLysProValValArgSerLeuSerLysA 2111
 : : : : :
 3017 CTTTCTCATGAC 3060
 2111 rgProIleGlyThrThrAspValAlaGluIleAsnAlaAspValTyr 2127
 : : : : :
 3061 AGATGTCGCGGAGCGCGCTGATCCGATCCGATCTCAAAACCCCTGCAT 3110
 2128 LeuCysMetThrGlnGluLysSerMetLysArgSerLeuLysG1 2144
 : : : : :
 3111 GCGAAGATCCTGACTTTTACCCAATCGGAT 3157
 2144 yLysGlyLysGluThrProValMetThrValHisGluAlaGlnGlyLysT 2161

3158 AAGAGG...TATTCAGATGTTCCACACTGTCATGAAGTCCAGGCGAGA 3204
2161 hrPheSerAspValLeuPheArgThrLysLysAlaAspSerLeu 2177
3205 CATACTCTGATGTTTCACTAGTAGTTAAGCCCTACACAGTCTCCATC 3254
2178 PheThr...LysGlnProHisIleLeuValGlyLeuSerArgHisThrAr 2193
3255 ATTGCAGGAGACGCCCATGTTTGGTCGTCATGTCACAGGACACCTG 3304
2193 gSerLeuValTyrAlaAla.....LeuSerSerLysLeuA 2205
3305 TTCGCTCAAGTACACTGTTGTTTATGATCCTTTAGTTAGTATCATTA 3354
2205 sPAsp.....LysValGlyThrTyrIleSerAsp 2214
3355 GAGATCTAGAGAACTTAGCTGCTACTTGTAGAT 3389

seq_name: /cgnl_7/ptodata/1/ina/6_COMB.seq:US-09-259-741-3

seq_documentation_block:
; Sequence 3, Application US/09259741
; Patent No. 6033895
; GENERAL INFORMATION:
; APPLICANT: GARGER, STEPHEN
; APPLICANT: HOLTZ, R. BARRY
; APPLICANT: MCCULLOCH, MICHAEL
; APPLICANT: TURPEN, THOMAS
; TITLE OF INVENTION: A PROCESS FOR ISOLATING AND
; TITLE OF INVENTION: PURIFYING VIRUSES SOLUBLE PROTEINS AND PEPTIDES FROM PLANT
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howrey & Simon
; STREET: 1299 Pennsylvania Avenue N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/259,741
; FILING DATE: February 25, 1999
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/037,751
; FILING DATE: March 10, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Halluin, Albert P
; REGISTRATION NUMBER: 25,277
; REFERENCE/DOCKET NUMBER: 00801.0140.US01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-463-8100
; TELEFAX: 650-463-8400
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6425 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: Genomic RNA
US-09-259-741-3

alignment_scores:
Quality: 278.00 Length: 1895
Ratio: 0.401 Gaps: 60

Percent Similarity: 36.570 Percent Identity: 16.570
alignment_block:
US-09-301-906-15 x US-09-259-741-3
Align seg 1/1 to: US-09-259-741-3 from: 1 to: 6425
382 ThrAlaIleValAlaGlyAspArgThrArgLeuGlyGlyLeuAl 398
9 ACAACAAUUAUCCAAACAAACAAACAAACAAACAAUUAUUAUUAU 58
398 aSerValAlaAsnAlaLeuAsnGlnGluValTyrSerSerValVal 415
59 UACAAUUAUUAUCCAAACAAACAAACAAACAAACAAUUAUUAUUA 108
415 eSerIleThrAsnArgLeuValLeuArgAspGlnSerAlaLeuLeu 431
109 ACACUGUCCGAGGAACAAUCCUUGGUCAUUAUGAUGCAAAAGCGUC 158
432 HisLeuAspThrLysLeuCysAspMetPheSerGlnArgAspAla 448
159 CUUUAACGACACAGCGGU...GAAGAGUUUAACGUCGUCAC..... 197
448 eArgGluLysProSerHisArgCysAspValPheLeuLysProArg 465
198 ...CGAGGCCCAAGGUAUUUUAUUAUUAUUAUUAUUAUUAUUA 243
465 rgGluLysLeuArgGluLeuPheProGluLeuSerIleGlnPheSer 481
244 CGCUUAUUAUCCAAACAAACAAACAAACAAACAAUUAUUAUUA 293
482 SerValArgSerHisPropheAlaAsnAlaMetArgSerCys... 497
294 ACGCAAAUUAUCCGAGGUAUCCUUGGAGGUAUUAUUAUUAUUA 343
497 eAsnGlyIlePheSerArgArgCysGlyAsnValCysPheAspIle 514
344 GGAUAUUAUUAUCCGAGGUAUUAUUAUUAUUAUUAUUAUUAU 393
514 lyGlySerPheThrTyrHisValLysAlaGlyHisValAsnCysHis 530
394 GCGGGAUUUAUCCGAGGUAUUAUUAUUAUUAUUAUUAUUAUUA 443
531 CysAsnProValLeuAspValLysAspValLysArgIleAsnGlu.. 546
444 UGCAUGCCCAACCGAGGUAUUAUUAUUAUUAUUAUUAUUAUUA 493
547IleLeuPheLeuSer.....ThrAlaGlyGly..... 555
494 AGACAGUAUUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 543
556AspSerTyr.....ValSerSer 561
544 UCCCAACUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 593
562 AspLeuLeuThrGluAlaSerLysSerValSerTyrCysSerArg 578
594 GAGCGUGUCUGACAAUUAUUAUUAUUAUUAUUAUUAUUAUUA 634
578 uSerGlnAsnCysAspSerArgAlaAspAlaGlyPheMetValAs 595
635 GCGGAGUCCAGCAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 684
595 yrAspIleSerProGlnValAlaAlaMetAspLysLysGlyAla 611
685 AUGACAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 734
612 LeuValPheAspIleAlaLeuMetPheProValGluLeuLeuTy 628
735 CAUACGUGCUAUGCCGCUUUAUUAUUAUUAUUAUUAUUAUUA 784
628 nGlyGluValTyrLeuGluLeuAspThrLeuValLysArgGlyAl 645

[illegible]

2154		:::	GUGUACAGGGUCGUAUAAGUU	2177
1546	ThrArgPheSerAsnAlaPheThrPheValAspAsnLeuLysGlyArgse	:::	UUUUAUCGAUAGCCUGGUAGCAUCACU	2218
2178	CAGCAAUAAGAAAC	:::	UUUUAUCGAUAGCCUGGUAGCAUCACU	2218
1562	rLaValPhePheSerLysLeuGlyGluGlyTyrThrTyraSnGlyGlys	:::::		1579
2219	AUCUGU	:::::		2225
1579	erHisValSerSerGlyTrpProArgAlaLeuGluAAspIleLeuThrAla	:::	GC GGUGUCGAUUCUGUCAAGAUC	1595
2226	IleLysTyrProSerValPheAspHisCysLeuValGlnLysTyrLysme	::::	CUUGAACCCGCUCAAAAGUU	2293
2250	CUCARAAGUACGUCUAUUGAC			1612
1612	tGlyGlyGlyValProPheHisAlaAspAspGluGlcystYrProSerA	::::		1629
2294	UGGA	::::		2297
1629	sPasnProIleLeuThrValasnLeuValGlyLysAlaasnPheSerThr			1645
2297	LysCysArgLysGlyGlyLysValMetValIleAsnValAlaLaserGlyAs	:::		1662
2298	pTyrPheLeuMetProCysGlyPheGlnArgThrHisLeuHisSerValA		GUCUUGGAUGUGGCAUCUACGAA	2320
1662	GUGUUUAUCAAAACCA			1679
2321	sSerIleAspGluGlyArgIleSerLeuThrPheArgAlaThrArgArg			1695
2336	ValPheGlyValGlyArgmetLeuGlnLeuAlaGlyGlyValSerAspGl			1712
2336	uLysSerProGlyValProAsnGlnGlnProGlnSerGlnGlyAlaThrA			1729
2336	rgThrIleThrProLysSerGlyGlyLysAlaLeuSerGluGlySerGly			1745
2337	ACGCCAAGAGAUCAUGCAUGGGGUGUUGAACCACCGCG			2378
1746	ArgGluValLysGlyArgSerThrTyrSerile		TrrPCy	1758
2379	AGGAAG		UAUCAUGUGGCGCUUUGGAUAUGA	2410
1758	sGluGlnAspTyrValArgLysCysGluTrpLeuArgAlaAspAsnProv			1775
2411	UCAGCAGGUGUGUGCAUGCAUGGAUGAUGGAGAGA			2447
1775	aIMetAlaLeuGluProAspTyrThrPrometThrPheGluValValLys			1791
2447	ThrGlyThrSerGluAspAlaValGluTyrLeuLysTyrLeuAlaIl			1808
2448	GUAGCUGUCAGCUCUGAGUGUGU		UAUUCGACAU	2485
1808	eGlyIleGluArgThrTyrArgAlaLeuMetAlaArgAsnIleAlay			1825
2486	GCGCAACUCAGAACUCUGCGACACUCUGCAACCGGAGAACCGCAUG			2535
1825	alThrAlaGluGlyValLeuLysValProAsnGlnValTyrGluSer			1841

```
2536 UCAGUAGCCAAAGGUUGUUCUU..... 2558
1842 LeuProGlyPheHisValTyrLysSerGlyThrAspLeuIlePheHis 1858
2558 ..... 2558
1858 rThrGlnAspGlyLeuArgValArgAspLeuProTyrValLeuIleAlaG 1875
2558 ..... 2558
1875 luLysGlyIlePheThrLysGlyLysAspValAspAlaValAlaLeu 1891
2559 .....GUGGCGGAGUU..... 2570
1892 GlyAspAsnLeuPheValCysAspIleLeuValPheHisAspAlaI 1908
2570 ..... 2570
1908 eAsnLeuIleGlyAlaLeuLysValAlaArgCysGlyMetValGlyGluS 1925
2570 ..... 2570
1925 erPheLysSerPheGluTyrLysCysTyrAsnAlaProProGlyGly 1941
2571 .....CCGGGCGUGGG 2582
1942 LysThrThrThrLeuValAspGluPheValLysSerProAsnSerThrAl 1958
2583 AAA.....ACCA 2590
1958 aThrIleThrAlaAsnValGlySerSerGluAspIleAsnMetAlaValL 1975
2591 AGAAAUUCUUCCAGGUGUUAUUUGAUGAAGAUUUUUUAGUACCG 2640
1975 ysLys.....ArgAspProAsnLeuGluGlyLeu 1984
2641 GGAACAGCCCGGAAAUAGUAGAGAGCGUGGAAUUCUCCAGGAU 2690
1985 AsnSerAlaThr.....ThrValAsnSerArgValValAs 1996
2691 AUUGUGCCACGACGACCAAGCUUAACCGUGUAUUCUUAUGAUGAA 2740
1996 nPhe.....IleValArgGlyMetTyrLysArgValLeuValAspGluV 2011
2741 UUUUGGAAAGACACACGCGUGUUAAGAGGUAUUAUUAUGAAG 2790
2011 aHisMetMetHisGlnGlnLeuGlnLeuGlyValPheAlaThrGly 2027
2791 GGUUGAUGUGCAUCUGUGUUAUUUUUUUUUGGCGCAUGUUAUG 2840
2028 AlaSerGluGlyLeuPhePheGlyAspIleAsnGlnIleProPheIleAs 2044
2841 UCGGAAAUUGCAUUGUUUACGGAGACACACAGCAGAUUCCAUACAUC 2890
2044 nArgGluLysValPheArgMetAspCysAlaValPheValProLysLysG 2061
2891 UAGAUUUUUGCAUUCUGUACCCGCGCCAUUUUUGCCAAUUGGAAG 2940
2061 luSerValValTyrThrSerLysSerTyrArgCysProLeuAspValCys 2077
2941 ACGAGUGGACACGACGACAGACUACUCCUUGUCCAGCGAUGUACACA 2990
2078 TyrLeuLeuSerSerMetThrValArgGlyThrGluLysCysTyrProG 2094
2991 CAUUUUCUGAAC.....ACGAGAUUGAGGG 3016
2094 uLysValValSerGlyLysAspLysProValValArgSerLeuSerLysA 2111
3017 CUUUGUCAUGAGC.....ACUUCUUGGUUAAAAAGUUCUGUUGCGCAG 3060
2111 rgProIleGlyThrThrAspAspValAlaGluIleAsnAlaAspValTyr 2127
3061 AGAUGGUGCGGGAGCCCGCGUGAUCUACCGAUCUACCAACCCUUGCAU 3110
```

```
2128 LeuCysMetThrGlnLeuGluLysSerAspMetLysArgSerLeuLysG 2144
3111 GGCAAGAUCCUGACUUUACCCCAUCCGAU...AAAGAAGCUCUGCUUC 3157
2144 yLysGlyLysGluThrProValMetThrValHisGluAlaGlnGlyLysT 2161
3158 AAGAGGG...UAUUCAGAUUGUUCACACUGUGCAUGAAGCAAGCGAGA 3204
2161 hrPheSerAspValValLeuPheArgThrLysLysAlaAspAspSerLeu 2177
3205 CAUACUCUGAUUGUUCACUAGUUAGUUAACCCUACACACGUCUCCAUC 3254
2178 PheThr...LysGlnProHisIleLeuValGlyLeuSerArgHisThrAr 2193
3255 AUUGCAGGAGACACGCCACAUUGUUUGGUGCAUUGUCAAAGGCACACCG 3304
2193 gSerLeuValTyrAlaAla.....LeuSerSerLysLeuA 2205
3305 UUCGCUCAAGUACACACUGUUGUUAUGAUCUUAUAGUAUCAUA 3354
2205 sAsp.....LysValGlyThrTyrIleSerAsp 2214
3355 GAGAUAGAGAAACUAGCUCGUAUUGUUAUAGU 3389
seq_name: /cgnl_7/ptodata/1/ina/6_COMB.seq:us-09-037-751-3
seq_documentation block:
; Sequence 3, Application US/09037751
; Patent No. 6037456
; GENERAL INFORMATION:
; APPLICANT: GARGER, STEPHEN
; APPLICANT: HOLTZ, R. BARRY
; APPLICANT: MCCULLOCH, MICHAEL
; APPLICANT: TURPEN, THOMAS
; TITLE OF INVENTION: A PROCESS FOR ISOLATING AND
; TITLE OF INVENTION: PURIFYING VIRUSES SOLUBLE PROTEINS AND PEPTIDES
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howrey & Simon
; STREET: 1299 Pennsylvania Avenue N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/037,751
; FILING DATE: 10-MAR-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Halluin, Albert P
; REGISTRATION NUMBER: 25,277
; REFERENCE/DOCKET NUMBER: 00801.0140.999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-463-8109
; TELEFAX: 650-463-8400
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6425 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: Genomic RNA
```

US-09-037-751-3

alignment_scores:		
Quality:	278.00	Length: 1895
Ratio:	0.401	Gaps: 60
Percent Similarity:	36.570	Percent Identity: 16.570

```

alignment_block:
US-09-301-906-15 x US-09-037-751-3  ..
Align seg 1/1 to: US-09-037-751-3 from: 1 to: 6425

382 ThrAlaIleValAlaGlyAspArgThrArgLeuGlyGlyClyLutIleLeuAl 398
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
9 ACAACAUAUACCAACAAACAACAACAACAACAACAACAUAUACAUAU 58
398 aserValAlaAsnAlaLeuAsnGlnGluValTyrSerSerValValAlS 415
||||:||||:||||:||||:||||:||||:||||:||||:||||:
59 UACAUAUACAAGGCAUACACACAGACAGCAUACCAUACAGCUUUGCG 108
415 erSerIleThrAsnArgLeuValLeuArgAspGlnSerAlaLeuLeuSer 431
||||:||||:||||:||||:||||:||||:||||:||||:||||:
109 ACACUGUCCGAGGAACAACUCCUUGGCUCAUAUGCAUACGAAAGCG 158
432 HisLeuAspThrLysLeuCysAspMetPheSerGlnArgAspAlaMetI 448
||||:||||:||||:||||:||||:||||:||||:||||:||||:
159 CUUUAACACACACAGCGGUU...GAAGAGUUAUACGCGCUGAC... 197
448 eArgGluLysProSerHisArgCysAspValPheLeuLysProArgGlu 465
||||:||||:||||:||||:||||:||||:||||:||||:||||:
198 ...CGCAGGCCCAAGGUAACUUUCAAAGUAUAUAGCGGAGGACGAC 243
465 rgGluLysLeuArgGluLeuPheProGluLeuSerIleGlnPheSerAs 481
||||:||||:||||:||||:||||:||||:||||:||||:||||:
244 CGCUUAUUGCUACCCGGCGGUAUCCAGAAUCCAAUUAUUAUUAUUA 293
482 SerValArgSerSerHisProPheAlaAsnAlaMetArgSerCys... 497
||||:||||:||||:||||:||||:||||:||||:||||:||||:
294 ACGCAAAUUGCGGUAUCCUUGCAGGUGGAUUGCAUUAUUAAGAUA 343
497 eAsnGlyIlePheSerArgArgCysGlyAsnValCysPhePheAspIle 514
||||:||||:||||:||||:||||:||||:||||:||||:||||:
344 GGAAUAUCUGAUGAUGCAAAUCCUACGGAUUAUUAUUAUUAUUA 393
514...LyGlySerPheThrTyrHisValLysAlaGlyHisValAsnCysH 530
||||:||||:||||:||||:||||:||||:||||:||||:||||:
394 CGCGGAUUUUUGCAUCCGCUUGCUUAGGAGCGAGCAUUAUGACAC 443
531 CysAsnProValLeuAspValLysAspValLysArgArgIleAsnGlu... 546
||||:||||:||||:||||:||||:||||:||||:||||:||||:
444 UGCAUGCCCAACUGGAGCUUUGCAGACAUCAUCCGCGCAGAGCCAG 493
547 .....IleLeuPheLeuSer.....ThrAlaGlyGly..... 555
||||:||||:||||:||||:||||:||||:||||:||||:||||:
494 AGACAGAUUUGCAUAUACCUUUGCAUAGGCUAGAGAGGGGGAACAG 543
556 .....AspSerTyr.....ValSerSer 561
||||:||||:||||:||||:||||:||||:||||:||||:||||:
544 UCCCAACAUUCCAAAGGAGCAUUGCAGAUACGCAAAUUAUCCGAA 593
562 AspLeuLeuThrGluAlaAlaSerLysSerValSerTyrCysSerArg 578
||||:||||:||||:||||:||||:||||:||||:||||:||||:
594 GACCGUGUCUUCACAUUUAUCCAGACAAUUGCGACAU.....CA 634
578 userGlnAsnCysAspSerArgAlaAspAlaGlyPheMetValAspVal 595
||||:||||:||||:||||:||||:||||:||||:||||:||||:
635 GCCGAUGCAGCAUACAGGCAGAGUGUAUGCCAUUGCGCUACACAGCA 684
595 yrAspIleSerProGlnValAlaGluAlaMetAspLysLysGlyAla 611
||||:||||:||||:||||:||||:||||:||||:||||:||||:
685 AUGACAUACCGGAGUAGUUGCGGCGCGACACUUAUGGAGAAAAAUG 734

```

[illegible]

1512 rLeuThrIleAsnGluGlyLysAlaGlyLysGlnLeuCysMetPheArgT 1529
:
2108 UAUGGCACAGCGAUAUCGUUAAUUCGAAGCAGAUGCUCUCAUU... 2153
1529 hrCySerCysGlyValGlnLeuAspValTyrrAsnGluAlaThrIleAla 1545
:
2154 GUGACACGGGCUAUUAAGUU 2177
1546 ThrArgPheSerAsnAlaPheThrPheValAspAsnLeuLysGlyArgSe 1562
:
2178 CAGCAAAUGAAAAC.....UUUAUCGAUAGCCUGGUAGCAUCACU 2218
1562 rAlavAlaPhePheSerLysLeuGlyGluGlyTyrrThrTyrrAsnGlyGlys 1579
:
2219 AUCUGCU..... 2225
1579 erHisValSerSerGlyTrpProArgAlaLeuGluAspileLeuThrAla 1595
:
2226 GCGUGUCGAUUCUGUCAAGAUC 2249
1596 IleLysTyrrProSerValPheAspHisCysLeuValGlnLysTyrrLysMe 1612
:
2250 CUCAAAGAUACAGCUGCUAUUGAC.....CUUGAAACC CGUCAAAAGUU 2293
1612 tGlyGlyGlyValProPheHisAlaAspAspGluGlyCysTyrrProSerA 1629
:
2294 UGGA..... 2297
1629 sPasnProIleLeuThrValAsnLeuValGlyLysAlaAsnPheSerThr 1645
2297 2297
1646 LysCysArgLysGlyLysValMetValIleAsnValAlaSerGlyAs 1662
:
2298GUCUUGGAUGUUGCAUCUAGGAA 2320
1662 pTyrrPheLeuMetProCysGlyPheGlnArgThrHisLeuHisSerValA 1679
:
2321 GUGUUUAUCAAACCA..... 2336
1679 snSerIleAspGluGlyArgIleSerLeuThrPheArgAlaThrArgArg 1695
2336 2336
1696 valPheGlyValGlyArgMetLeuClnLeuAlaGlyLysValSerAspGl 1712
2336 2336
1712 uLysSerProGlyValProAsnGlnGlnProGlnSerGlnGlyAlaThra 1729
2336 2336
1729 rgThrIleThrProLysSerGlyLysAlaLeuSerGluGlySerGly 1745
||| ||||| : : : : : : : : : : : : : : : : : : :
2337ACGCCCAAGAGCAUGCAUGGGUGUGUUGAANACCACCGCG 2378
1746 ArgGluValLysGlyArgSerThrTyrrSerIle.....TrpCy 1758
||| :
2379 AGGAAG.....UAUCAUGUGCGCUUUGGAAUAUGA 2410
1758 sGluGlnAspTyrrValargLysCysGluTrpLeuArgAlaaspAsnProv 1775
||||||| ||| : : : : : : : : : : : : : : : : : : :
2411 UGACAGCGGUGUGGUGACAUCGCAUGAUUGGAGAAGA..... 2447
1775 alMetaIlaLeuGluProAspTyrrThrPrometThrPheGluValValLys 1791
2447 2447
1792 ThrGlyThrSerGluAspAlaValValGluTyrrLeuLysTyrrLeuAlaIl 1808
:
2448 GUAGCUGCACUCUGAGUCUGUGUU.....UAUUCGCGACAU 2485

1808	eGlyIleGluArgThrTyrArgAlaLeuLeuMetAlaArgAsnIleAlaVal	1825
2486	GGCGAAACUCAGAACUCUGCGCAGACUGCUCCGAAACGGAGAACCGCAUG	2535
1825	alThrThrAlaGluGlyValLeuLeuLysValProAsnGlnValTyrGluSer	1841
2536	UCAGUAGCGCAAGAGGUUGUUCUU	2558
1842	LeuProGlyPheHisValTyrLysSerGlyThrAspLeuIlePheHis	1858
2558		2558
1858	rThrGlnAspGlyLeuArgValArgAspLeuProTyrValLeuIleAlaG	1875
2558		2558
1875	IuLysGlyIlePheThrLysGlyLysAspValAspAlaValAlaLeu	1891
2559		2570
1892	GlyAspAsnLeuPheValCysAspAlaLeuValPheHisAspAlaIle	1908
2570		2570
1908	eAsnLeuIleGlyAlaLeuLysValAlaArgCysGlyMetValGlyGlu	1925
2570		2570
1925	erPheLysSerPheGluTyrLysCysTyrAsnAlaProProGlyGlyGly	1941
2571		2582
1942	LysThrThrLeuValAspGluPheValLysSerProAsnSerThrAl	1958
2583	AAA	2590
1958	aThrIleThrAlaAsnValGlySerSerGluAspIleAsnMetAlaVal	1975
2591	AGAAAUUCUCCAGGGUUAAUUUGAUGAAGAUUUUAGUACCU	2640
1975	ysLys	1984
2641	GGAAGCAAGCCGGAAUGAUCAGAAAGACGUGCGAAUUCUCCAGGAU	2690
1985	AsnSerAlaThr	1996
2691	AUUGUGCCCGACGAGCAACCGUUAAACCGUUGAUUUCUUGAUGAA	2740
1996	nPhe	2011
2741	UUUUGGAAAAAGCACACGUCUGACUUAAGAGGUUUUAUUAUGAUGAAG	2790
2011	alHisMetMetHisGlnGlyLeuLeuGlnLeuGlyValPheAlaThrGly	2027
2791	GGUUGAUGUUGCAUCACUGGUGUUUUUUUUGUGCGGAGUUGCAUUG	2840
2028	AlaSerGluGlyLeuPhePheGlyAspIleAsnGlnIleProPheIleAs	2044
2841	UGCGAAUUGCAUUGUUUACGGAGACACACAGCAGAUUCCAUACACAA	2890
2044	nArgGluLysValPheArgMetAspCysAlaValPheValProLysLysG	2061
2891	UAGAGUUUCAGGAUUCCGUACCCGCCCAUUUGCCAAUUGGAAGUUG	2940
2061	IuSerValValTyrThrSerLysSerTyrArgCysProLeuAspValCys	2077
2941	ACGAGGUGGAGACACGCAAGAACUACUCUGGUGCCAGCGAUGUACCA	2990
2078	TyrLeuLeuSerSerMetThrValArgGlyThrGluLysCysTyrProG	2094
2991	CAUUAUCUAGAC	3016
2094	uLysValValSerGlyLysAspLysProValValArgSerLeuSerLysA	2111


```

; REFERENCE/DOCKET NUMBER: 00801.0140.9999
;
; TELECOMMUNICATION INFORMATION:
;
; TELEPHONE: 650-463-8109
;
; TELEFAX: 650-463-8400
;
; TELEX:

```

alignment_scores:		
Quality:	278.00	Length: 1895
Ratio:	0.401	Gaps: 60
Percent Similarity:	36.570	Percent Identity: 16.570

alignment_block:

US-09-301-906-15 x US-09-037-751-5

Align seg 1/1 to: US-09-037-751-5 from: 1 to: 6446

```

382 ThrAlaIleValAlaGlyAspArgThrArgLeuGlyGlyGluLeuAl 398
|||:||||:||||:||||:||||:||||:||||:||||:||||:
9 ACAACAUAUCCAAACAACAACAACAACAACAACAUAUAUAUAU 58
|||:||||:||||:||||:||||:||||:||||:||||:||||:
398 aSerValAlaAsnAlaLeuAsnGlnGluValTyrSerSerValVal 415
|||:||||:||||:||||:||||:||||:||||:||||:||||:
59 UACAUAUAACAUGCAUACACACACAGACAGCUACACAUCAUUGCU 108
|||:||||:||||:||||:||||:||||:||||:||||:||||:
415 eRserIleThrAsnArgLeuValLeuArgAspGlnSerAlaLeuTyr 431
|||:||||:||||:||||:||||:||||:||||:||||:||||:
109 ACACUGUCCGAGGAACAACUCCUUGGUCAUGAUGCAUACAAGCGUC 158
|||:||||:||||:||||:||||:||||:||||:||||:||||:
432 HisLeuAspThrLysLeuCysAspMetPheSerGlnArgAspAlaMe 448
|||||:||||:||||:||||:||||:||||:||||:||||:
159 CUUUACGCACACGCGGUU...GAAGAUUUUAACCCUCCUGAC... 197
|||:||||:||||:||||:||||:||||:||||:||||:||||:
448 eArgGluLysProSerHisArgCysAspValPheLeuTyrProArgL 465
|||:||||:||||:||||:||||:||||:||||:||||:||||:
198 ...CGCAGGCCCAAGGUGAACUUUCAAAGAAUAUAAGCGAGGAGCA 243
|||:||||:||||:||||:||||:||||:||||:||||:||||:
465 rGluLysLeuArgGluLeuPheProGluLeuSerIleGlnPheSer 481
|||:||||:||||:||||:||||:||||:||||:||||:||||:
244 CGUUUAUUGCUACCCGGGGUUAUCCAGAAUUCCAAUUUAUAUUUA 293
|||:||||:||||:||||:||||:||||:||||:||||:||||:
482 SerValArgSerHisProPheAlaAsnAlaMetArgSerCys... 497
|||:||||:||||:||||:||||:||||:||||:||||:||||:
294 ACGCAAAAUUGCCGUGCAUUCGCUUGCAGGUGAUUGCAUUAUAAG 343
|||:||||:||||:||||:||||:||||:||||:||||:||||:
497 eAsnGlyIlePheSerArgArgCysGlyAsnValCysPhePheAsp 514
|||:||||:||||:||||:||||:||||:||||:||||:||||:
344 GGAUAUUCGAUGAUGCAAAUCCCUACGGGAUCAUUGCAUUAUGAC 393
|||:||||:||||:||||:||||:||||:||||:||||:||||:
514 lyGlySerPheThrTyrHisValLysAlaGlyHisValAsnCysHis 530
|||:||||:||||:||||:||||:||||:||||:||||:||||:
394 GCGGGAUUUUGCAUUGCAUCUGUUAUAGGACGAGCAUUGACACUG 443
|||:||||:||||:||||:||||:||||:||||:||||:||||:
531 CysAsnProValLeuAspValLysAspValLysArgIleAsnGlu... 546
|||:||||:||||:||||:||||:||||:||||:||||:||||:
444 UGAUGGCCCAACUUGGAGCUUUCGAGCAUCAUGCGGCACGAAGGCC 493
|||:||||:||||:||||:||||:||||:||||:||||:||||:
547 .....IleLeuPheLeuSer.....ThrAlaGlyGly..... 555
|||||:||||:||||:||||:||||:||||:||||:||||:
494 AGACAGUAUUGAUAUAUCCUUUACGGCUNAGACAGAGGGGGGAAAC 543
|||:||||:||||:||||:||||:||||:||||:||||:||||:
556 .....AspSerTyr.....ValSerSer 561
|||:||||:||||:||||:||||:||||:||||:||||:||||:
544 UCCCCAACUUCAAAGGAAGCAUUGGAGCAUAGCAUAGCAUUAUUG 593
|||:||||:||||:||||:||||:||||:||||:||||:||||:

```

1161 UGGUUUCCCAAAUGAGGGAUUGUCAUCCACCA 1196
875 gLysValProGlySerValValValThrIleCysThrSerGlyAlaSerA 892
1196 1196
892 sPArgLeuGluLeuArgGlyAlaPheAspIleSerLysGluThrPheGly 908
1197UUAUUCGACAUUUCUUGGAGACUAGUAA 1226
909 ArgLysLeuLysAsnSerArgLeuArgValPheSerArgAl 922
1227 AGGACGCCAAGGAAGUUCUAGUUCUCCAGGAUUUCGUUUUACA 1271
922 aIleValGluAspSerIleLysValMetLysAlaMetLysThrGluAspG 939
1272GUGCUUAAACCACAUUCGAACAUACCAGG 1299
939 lYlsProLeuProIleThrGluAspSerValTyrAlaPheIleMetGly 955
1300 CGAAAGCUCUUACAUACGCAAAUGUUUUGUCCUUUGCGAAUCG 1343
956 AsnValSerAsnValHisCysThrArgAlaGlyLeuLeuGlyGlySerLy 972
1344 AUUCGACUGGAGGUAAUCAUUUACGGUGUGAC 1375
972 sAlaThrValValSerSerValSerLysGlyLeuValAlaArgGlyAlaA 989
1376 AGCGAGGUCCGAUUGGAGUGGCAAAUUCUUUUAACA 1415
989 laThrLysAlaPheSerGlyIleThrSerPheSerThrGlySerLeu 1005
1416UCCUUGUCCAUUGACG 1430
1006 PheTyrAspArgGlyLeuThrGluAspGluArgLeuAspAlaLeuValAr 1022
1431 UUUUAC 1436
1022 gThrGluAsnAlaIleAsnSerProValGlyIleLeuGluThrSerArgV 1039
1437CUGCAUACUAGCUUGCCGUUUAAGGAUACU 1470
1039 aAlaValSerLysValValAlaGlyThrLysGluPhe 1051
1471 UACUGAUUAGCAAGUUUAGUCUGGUUCGAAACGGUGGCCAGCAGUG 1520
1052 TrpSerGluValSerLeuAsnAspPheThrThrPheValLeuArgAsnLy 1068
1521 UGGGAUGAGAUUUCGUG 1538
1068 sValLeuIleGlyPheValAlaSerLeuGlyAlaAlaProIleAlaT 1085
1538 1538
1085 rPlysTyrArgArgGlyIleAlaAlaAsnAlaArgArgTyrAlaGlySer 1101
1538 1538
1102 SerTyrGluThrLeuSerSerLeuSerSerGlnAlaAlaGlyGlyLeuAr 1118
1538 1538
1118 gGlyLeuThrSerSerThrValSerGlyGlySerLeuValValArgArg 1135
1538 1538
1135 lYPheSerSerAlaValThrValThrArgAlaThrValAlaLysArgGln 1151
1538 1538
1152 ValProLeuAlaLeuLeuSerPheSerThrSerTyrAlaIleSerGlyCy 1168

1538 1538
1168 sSerMetLeuGlyIleTrpAlaHisAlaLeuProArgHisLeuMetPheP 1185
1538 1538
1185 hePheGlyLeuGlyThrLeuLeuGlyAlaArgAlaSerAlaAsnThrTrp 1201
1538 1538
1202 LysPheGlyGlyPheSerAsnAsnTrpCysAlaValProGluValValTr 1218
1539CGUUUUGGAACGCAUUUCCUCCUGGAAAGAGAGCGCUU 1579
1218 pArgGlyLysSerValSerSerLeuLeuLeuProIleThrLeuGlyValS 1235
1580 GAACAGGAACUUUACAGAGUGCGAGCGCAGC 1611
1235 erLeuIleIleArgGlyLeuLeuAsnAspThrIleProGlnLeuAlaTyr 1251
1612 CAUUAGAGAUACAGGUGCCUGAUCUA...UAU 1640
1252 ValProProValGluGlyArgAsnValTyrAspGluThrLeuArgTyrTy 1268
1641 GUG 1643
1268 rArgAspPheAspTyrAspGluGlyAlaGlyProSerGlyThrGlnHisG 1285
1644ACCUCCACG 1653
1285 luAlaValProGlyAspAspAsnAspGlySerThrSerValSerSer 1301
1654 ACAGAUUAGUGACUGAG 1670
1302 TyrAspValValThrAsnValArgAspValGlyIleSerThrAsnGlyGl 1318
1671 UACAGCCUUGUGGACAUUGCCUGGACAUUAGGAAGAAGAUUGGA 1720
1318 uValThrGlyGluGluGluThrHisSerProArgSerValGlnTyrThrT 1335
1721 AGAAACGGAUGUGAUGUACAAUG 1743
1335 yrValGluGluGluValAlaProSerAlaValAlaGluArgGlnGly 1351
1744 CACUUUCAGAG 1754
1352 AspProSerGlySerGlyThrAlaAspAlaMetAlaPheValGluSerVa 1368
1755UUAUCGGUGUUUAGGAGUC 1774
1368 lLysLysGlyValAspValPheHisGlnSerSerGlyGluThrA 1385
1775 UGACAAAUUCGAUGUUUUAUUUCCAGAUU 1809
1385 laArgGluValGluValAspGlyLysGlyLeuLeuProGluSerValVal 1401
1810 GCCAAUUCUUUGGAAGUUGACCCAAUGACGCGCAGGAAGUUUA 1853
1402 GlyGluAlaProThrGlnGluArgGlyArgAl 1412
1854 GUGCGGUCAUGAGCAUAGAGCGGUGUCACUCACAUUUUGAACGCC 1903
1412 aAlaAspGlyAsnThrAlaGlnThrAlaValAsnGluGlyAspArgGluP 1429
1904 UACUGAGCGCAUUGUUGCGCUAGCUUACAGGAUCAAGAGAGCGCU 1950
1429 roValGlnSerSerLeuValSerSerProGlnAlaAspIleProLysVal 1445
1951 CAGAAAGGUGCUUUGGUAGUUAUCCUACAAGAGAAGUUAGAAGCCGCCAUG 2000
1446 ThrGlnSerGluValHisAlaGlnLysGluValLysGlnGluValProLe 1462
2001 AAGGCUUCGAUGCCACAGAGAGAUUACAUAU 2032


```
||||| ... ||||| ... ||||| ...
2891 UAGAGUUCAGAGUCCCGACCCCAUUUUGCCAAUUGGAAGUUG 2940
2061 luSerValValTyrThrSerLysSerTyrArgCysProLeuAspValCys 2077
...:||||| ...: ||||| |||||
2941 ACAGGUGGAGACACGACAGACUACUCUCCGUUGCCAGCCGAGUACACA 2990
2078 TyrLeuLeuSerSerMetThrValArgGlyThrGluLysCysTyrProG 2094
...: ||||| ...: |||||
2991 CAUUAUCUGAAC.....AGGAGUAUGAGGG 3016
2094 uLysValValSerGlyLysAspLysProValValArgSerLeuSerLysA 2111
|||||:||||| ...: ||||| ...: |||||:|||||:
3017 CUUUGUCAUGAC.....ACUUCUGGUAUAAAGUCUGUUCGCCAGG 3060
2111 rgProTleGlyThrThrAspValAlaGluLeuAsnAlaAspValTyr 2127
...: ||||| ...: |||||
3061 AGAUGGUGCGGCGCGCGGUGAUAUCCGGAUACUCAAACCCUUGCAU 3110
2128 LeuCysMetThrGlnLeuLysSerAspMetLysArgSerLeuLysG 2144
...: |||||:||||| |||||:||||| ...: |||||:
3111 GCAAGAUCCUGACUUAUCCCAUUGCGAU...AAGAAGCUCUGCUUUC 3157
2144 yLysGlyLysGluThrProValMetThrValHisGluAlaGlnGlyLysT 2161
...: ||||| ...: ||||| ||||| ||||| ||||| ||||| |||||
3158 AGAGGG...UAUUCAGUUGUACACUGUGCAUGAUGGCAAGGCGGAGA 3204
2161 hrPheSerAspValValLeuPheArgThrLysLysAlaAspSerLeu 2177
|||||:||||| ||||| ||||| ...: |||||
3205 CAUACUCUGAUUUCACUAGUAGGUUAAGGUAAACCCUACACACGCUCCAUC 3254
2178 PheThr...LysGlnProHisLeuValGlyLeuSerArgHisThrAr 2193
...: ||||| ...: ||||| ||||| ||||| ||||| ||||| |||||
3255 AUUGCAGGAGACAGCCCAUGUUGGUGCAUGUUGUAGGACACACAC 3304
2193 gSerLeuValTyrAlaAla.....LeuSerSerLysLeuA 2205
||||| ||||| ...: ||||| ||||| |||||
3305 UUGGCUAAGUACUACACUGUUGUUAUGAUCUUAUGUUAUGUUAUGAUA 3354-
2205 spAsp.....LysValGlyThrTyrIleSerAsp 2214
||| |||||:|||||: ||||| |||||
3355 GAGAUACAGAGAAACUACGUCGUAUUGUUAUGAU 3389
```

seq_name: /cgnl_7/ptodata/1/ina/6_COMB.seq:us-09-259-741-4

```
seq_documentation_block:
; Sequence 4, Application US/09259741
; Patent No. 6033895
; GENERAL INFORMATION:
; APPLICANT: GARGER, STEPHEN
; APPLICANT: HOLTZ, R. BARRY
; APPLICANT: MCCULLACH, MICHAEL
; APPLICANT: TURPEN, THOMAS
; TITLE OF INVENTION: A PROCESS FOR ISOLATING AND
; TITLE OF INVENTION: PURIFYING VIRUSES SOLUBLE PROTEINS AND PEPTIDES FROM PLANT
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howrey & Simon
; STREET: 1299 Pennsylvania Avenue N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/259,741
; FILING DATE: February 25, 1999
; CLASSIFICATION:
```

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/037,751
; FILING DATE: March 10, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Halluin, Albert P
; REGISTRATION NUMBER: 25,277
; REFERENCE/DOCKET NUMBER: 00801.0140.US01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-463-8100
; TELEFAX: 650-463-8400
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6475 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: Genomic RNA
; US-09-259-741-4

alignment_scores:
Quality: 278.00 Length: 1895
Ratio: 0.401 Gaps: 60
Percent Similarity: 36.570 Percent Identity: 16.570

alignment_block:
US-09-301-906-15 x US-09-259-741-4 ..
Align seg 1/1 to: US-09-259-741-4 from: 1 to: 6475

382 ThrAlaIleValAlaGlyAspArgThrArgLeuGlyGlyGluLeuAl 398
|||||:||||| ...: |||||
9 ACACAAUUACCAACAAACAACAACAACAACAACAACAUAUACAUAUACU 58
398 aserValAlaAsnAlaLeuAsnGlnGluValTyrSerSerValValS 415
...: ||||| ||||| ||||| ||||| ||||| |||||
59 UCAAAUUAUACGAGGCAUACACAGCAGCAGCAGCAGCAGCAGCAGCAG 108
415 erSerIleThrAsnArgLeuValLeuArgAspGlnSerAlaLeuLeuSer 431
...: ||||| ||||| ||||| ||||| |||||
109 ACACUGUCGAGGAGAAACAACUCCUUGGCAUAGUACUAGCAAGCGUGU 158
432 HisLeuAspThrLysLeuCysAspMetPheSerGlnArgAspAlaMetI 448
||||| ||||| ...: |||||
159 CUUUAACGACACAGCGGU...GAAGAGUUUAACGCGUGAC..... 197
448 eArgGluLysProSerHisArgCysAspValPheLeuLysProArgGluA 465
...: ||||| ...: |||||
198 ...CGCAGGCCCAAGGUGAAGUUUCAAAGUAAUAGCGAGGAGCAGA 243
465 rgGluLysLeuArgGluLeuPheProGluLeuSerIleGlnPheSerAsp 481
...: ||||| ||||| ||||| ||||| ||||| |||||
244 CGCUUAUUCUACCCGGCGGUUACUCCAGAAUCCAAAUUAUUAUUAUAAC 293
482 SerValArgSerSerHisProPheAlaAsnAlaMetArgSerCys...Ph 497
...: ||||| ||||| ||||| ||||| ||||| |||||
294 ACCGAAAUAUGCCGUGCAUUCGUGGAGGUGGAGUUCGGAUUAUUAACU 343
497 eAsnGlyIlePheSerArgArgCysGlyAsnValCysPheAspIleG 514
...: ||||| ...: |||||
344 GGAUAUACUGAUGCAUAAUCCUACGGAUCAUAGACUUAUAGCAUAG 393
514 lyGlySerPheThrTyrHisValAlaGlyHisValAsnCysHisVal 530
|||||:|||||: |||||: |||||: |||||: |||||: |||||
394 GCGGGAUUUUUGCAUGCGCAUCUGUUAAGGAGGAGGAGCAUUAUAGUAC 443
531 CysAsnProValLeuAspValLysAspValLysArgArgIleAsnGlu... 546
||| ||| ||||| ||||| ||||| ||||| ||||| |||||
444 UGCAUGCCCAACCGGACCGUCCGAGACAUACUGCGGACGAGCCGAGAA 493
547 .....IleLeuPheLeuSer.....ThrAlaGlyGly..... 555
|||||:|||||
```


1135 lyPheSerAlaValThrValThrArgAlaThrValAlaLysArgGln 1151
1538
1152 valProLeuAlaLeuLeuSerPheSerThrSerTyrAlaIleSerGlyCy 1168
1538
1168 sSerMetLeuGlyIleTrpAlaHisAlaLeuProArgHisLeuMetPheP 1185
1538
1185 hepGlyLeuGlyThrLeuLeuGlyAlaArgAlaSerAlaAsnThrTrp 1201
1538
1202 LysPheGlyGlyPheSerAsnAsnTrpCysAlaValProGluValValTr 1218
1539
1218 parGlyLysSerValSerSerLeuLeuLeuProIleThrLeuGlyValS 1235
1580 GAACAGGAACAUUCAGAG.....GUGGACGAGCGACG 1611
1235 erLeuIleIleArgGlyLeuLeuAsnAspThrIleProGlnLeuAlaTyr 1251
1612 CAUUGAGAGCAAG.....GUGCCUGAUCUA...UAU 1640
1252 ValProProValIleGlyArgAsnValTyrAspGluThrLeuArgTyrTy 1268
1641 GUG.....
1268 rArgaspPheAspTyrAspGluGlyAlaGlyProSerGlyThrGlnHisG 1285
1644
1285 luAlaValProGlyAspAspAspGlySerThrSerSerValSerSer 1301
1654 ACAGAUUA.....GUGACUGAG 1670
1302 TyrAspValValThrAsnValArgAspValGlyIleSerThrAsnGlyGI 1318
1671 UACAAGGCCUGUGGACAUCCUGCGCUUACAUUAGGAAGAUGCA 1720
1318 uValThrGlyGluGluThrHisSerProArgSerValGlnTyrThrT 1335
1721 AGAAACG.....GAAGUGAUGUACAAG 1743
1335 yCValGluGluGluValAlaProSerAlaAlaValAlaGluArgGlnGly 1351
1744 CACUUCACAG..... 1754
1352 AspProSerGlySerGlyThrAlaAspAlaMetAlaPheValIleSerVa 1368
1755UUAUCGUGUUAAGGAGUC 1774
1368 lLysLysGlyValAspAspValPheHisGlnGlnSerSerGlyGluThra 1385
1775 UGACAAAUUCGAUGUUGAUUUUUUCCAGAG.....U 1809
1385 laArgGluValGluValAspGlyLysGlyLeuLeuProGluSerValVal 1401
1810 GCCAUCUUUGGAAGUUGAC.....CCAAUGACGCGACGGAAGGUUAU 1853
1402 GlyGluAlaProThrGlnGluArgGly.....ArgAl 1412
1854 GUGCCGUCUACGACAAUGAGCGGUGUCUCACAUUUGAACGACC 1903
1412 aAlaAspGlyAsnThrAlaGlnThrAlaValAsnGluGlyAspArgGluP 1429
1904 UACUGAGGCGAAUGUUGCG...CUAGCUUUACAGGAUCAAGAGGCUU 1950
1429 roValGlnSerSerLeuValSerSerProGlnAlaAspIleProLysVal 1445

1951 CAGAAAGGUCUUUGUAGUACCUCAAGAGAGAUUGAAGAACCCUCCAU 2000
1446 ThrGlnSerGluValHisAlaGlnLysGluValLysGlnGlnValProLe 1462
2001 AAGGGUUCG.....AUGGCCAGAGGAGAUUACAUAU 2032
1462 uAlaThrValSerGlyAlaThrProIleValAspGluLysProAlaProS 1479
2033 AGCUGGUCUUGGAGAUCAUCG..... 2057
1479 erValThrThrArgGlyValLysIleIleAspLysGlyLysAlaValAla 1495
2057 2057
1496 HisValAlaGluLysLysGlnValGlnValGluGlnProLysGlnArgse 1512
2058 GAGUCGUCCUAUUAAGAAGAGGAGAGUAGUCUUUAGAGCAGUUUCA 2107
1512 rLeuThrIleAsnGluGlyLysAlaGlyLysGlnLeuCysMetPheArgT 1529
2108 UAUGGCAACGGCAGAUUCGUAAUUCGUAAGCAGAGCUCGAU... 2153
1529 hrCysSerCysGlyValGlnLeuAspValTyrAsnGluAlaThrIleAla 1545
2154GUGUACACGGGUCGCGAUUAAGUU 2177
1546 ThrArgPheSerAsnAlaPheThrPheValAspAsnLeuLysGlyArgse 1562
2178 CAGCAAAUGAAAAAC.....UUUUCGAUAGCCUGGUAGCAUCACU 2218
1562 rAlaValPhePheSerLysLeuGlyGluGlyTyrThrTyrAsnGlyGlys 1579
2219 AUCUGCU..... 2225
1579 erHisValSerSerGlyTyrProArgAlaLeuGluAspIleLeuThrAla 1595
2226GCGUGUCGAAUCUCGUCAGAU 2249
1596 IleLysTyrProSerValPheAspHisCysLeuValGlnLysTyrLysme 1612
2250 CUCAAAAGUACAGCUCUAUUGAC.....CUUGAAACCCGUCAAAGUU 2293
1612 tGlyGlyGlyValProPheHisAlaAspAspGluGluCysTyrProSera 1629
2294 UGGA..... 2297
1629 spAsnProIleLeuThrValAsnLeuValGlyLysAlaAsnPheSerThr 1645
2297 2297
1646 LysCysArgLysGlyGlyLysValMetValIleAsnValAlaSerGlyAs 1662
2298GUCUUGGAUGUUGCAUCUAGGAA 2320
1662 pTyrPheLeuMetProCysGlyPheGlnArgThrHisLeuHisSerVala 1679
2321 GUGGUUAUCAAAACCA..... 2336
1679 snSerIleAspGluGlyArgIleSerLeuThrPheArgAlaThrArgArg 1695
2336 2336
1696 ValPheGlyValGlyArgMetLeuGlnLeuAlaGlyGlyValSerAspGI 1712
2336 2336
1712 uLysSerProGlyValProAsnGlnGlnProGlnSerGlnGlyAlaThra 1729
2336 2336
1729 rgThrIleThrProLysSerGlyGlyLysAlaLeuSerGluGlySerGly 1745
1745

```
2337 .....ACGCCAAGAGUCAUGCAUGGGUGUUGUAGAAACCCACGCG 2378
1746 ArgGluValLysGlyArgSerThrTyrSerIle.....TrpCys 1758
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2379 AGGAAG.....UAUCAGUGGGCGCUUUGGAUAUGA 2410
1758 sGluClnAspTyrValArgLysCysGluTrpLeuArgAlaAspAsnProv 1775
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2411 UGAGCAGGGUGUGGACAGUGCAUGGAUUGGAGAAGA..... 2447
1775 aMetAlaLeuClnProAspTyrThrProMetThrPheGluValValLys 1791
2447 ..... 2447
1792 ThrGlyThrSerGluAspAlaValValGluTyrLeuLysTyrLeuAlaIle 1808
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2448 GUAGGUGUCAGCUCUGAGCUGUGUUGU.....UAUCCGACAU 2485
1808 cGlyIleGluArgThrTyrArgAlaLeuLeuMetAlaArgAsnIleAlaVal 1825
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2486 GCGAAGACUCAGACUCUGCGCAGACUCUGCGAAGCGGAGAACCGCAUG 2535
1825 alThrThrAlaClnGluValLeuLysValProAsnGlnValTyrClnSer 1841
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2536 UCAGUAGCGCAAGAGGUGUGUCUU..... 2558
1842 LeuProGlyPheHisValTyrLysSerGlyThrAspLeuIlePheHis 1858
2558 ..... 2558
1858 rThrClnAspGlyLeuArgValArgAspLeuProTyrValLeuIleAlaG 1875
2558 ..... 2558
1875 luLysGlyIlePheThrLysGlyLysAspValAspAlaValAlaLeu 1891
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2559 .....GUGGAGCGGAGU..... 2570
1892 GlyAspAsnLeuPheValCysAspAlaLeuValPheHisAspAlaIle 1908
2570 ..... 2570
1908 eAsnLeuIleClyAlaLeuLysValAlaArgCysGlyMetValGlyGlu 1925
2570 ..... 2570
1925 erPheLysSerPheClnTyrLysCysTyrAsnAlaProProGlyGlyGly 1941
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2571 .....CCGGCGGUGGG 2582
1942 LysThrThrThrLeuValAspGluPheValLysSerProAsnSerThrAl 1958
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2583 AAA.....ACCAA 2590
1958 aThrIleThrAlaAsnValGlySerSerGluAspIleAsnMetAlaVal 1975
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2591 AGAAAUUCUCCAGGGUAAUUUUGAAGAAAGUAAUUUUGAUGACCG 2640
1975 ysLys.....ArgAspProAsnLeuGluGlyLeu 1984
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2641 GGAACAGACCGCGAAUGAUCAGAACGGUGCGAAUUCUCAGGGAU 2690
1985 AsnSerAlaThr.....ThrValAsnSerArgValValAs 1996
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2691 AUUGGGCCAGGAGGACAGUUAARACCGUAGUUAUUCUUGAUGAA 2740
1996 nPhe.....IleValArgGlyMetTyrLysArgValLeuValAspGlu 2011
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2741 UUUUGGAAAGACACACCGCUGAGUUAAGAGGUUAUUCUUGAUGAAG 2790
2011 alHisMetMetHisClnGlyLeuLeuGlnLeuGlyValPheAlaThrGly 2027
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2791 GGUUGAUGUUGCAUCAGGGUGUGUAAUUUUGGCGGCAUGCAUUG 2840
```

```
2028 AlaSerGluGlyLeuPhePheGlyAspIleAsnGlnIleProPheIleAs 2044
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2841 UCGAUAUUGCAUUGUUUUGCGAGACACAGCAGAUUCCAUCAUCA 2890
2044 nArgGluLysValPheArgMetAspCysAlaValPheValProLysLysG 2061
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2891 UAGAUUUUCAGGAUUCGCGACCGCCCAUUUUGCCAAAUUGGAAGUUG 2940
2061 luSerValValTyrThrSerLysSerTyrArgCysProLeuAspValCys 2077
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2941 ACGAGGUGGAGACAGCAGAAACUCCUGUGUCCAGCGCAUGUCACA 2990
2078 TyrLeuLeuSerSerMetThrValArgGlyThrGluLysCysTyrProG 2094
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2991 CAUUAUCUGAAC.....AGGAGUAUGAGGG 3016
2094 ulysValValSerGlyLysAspLysProValValArgSerLeuSerLys 2111
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3017 CUUUGUCAUGAGC.....ACUUCUUGGUUAAAAAGUCUGUUGCGAGG 3060
2111 rgProIleGlyThrThrAspAspValAlaGluIleAsnAlaAspValTyr 2127
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3061 AGAUGGUGCGGCGCGCCCGUGAUCAUCCGCAUCUCAAACCCUUGCAU 3110
2128 LeuCysMetThrGlnLeuGluLysSerAspMetLysArgSerLeuLysG 2144
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3111 GCGAAGAUCCUGACUUUUAUCCCAUCCGAU.....AAAGAAGCUCUGCUUUC 3157
2144 yLysGlyLysGluThrProValMetThrValHisGluAlaGlnGlyLys 2161
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3158 AAGAGGG...UAUUCAGAUUGUUCACACUGUGCAUGAAGUGCAAGCGAGA 3204
2161 hrPheSerAspValValLeuPheArgThrLysLysAlaAspAspSerLeu 2177
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3205 CAUACUCUGAUUGUUCACUAGUAGGUUAACCCUACACACAGUCUCCAUC 3254
2178 PheThr...LysGlnProHisIleLeuValGlyLeuSerArgHisThrAr 2193
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3255 AUUGCAGGAGACAGCCACAUUGUUUGGUGCAUUGUCAAGGCACACCG 3304
2193 gSerLeuValTyrAlaAla.....LeuSerSerLysLeuA 2205
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3305 UUCGCUCAAGUACUACACUGUUGUUAUGAUCCUUUAGUAGUAUAUA 3354
2205 sPAsp.....LysValGlyThrTyrIleSerAsp 2214
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3355 GAGAUCAAGAAACUUAAGCUGUACUUGUAGAU 3389
seq_name: /cgnl_7/ptodata/1/ina/6_COMH.seq:US-09-037-751-4
seq_documentation_block:
; Sequence 4; Application US/09037751
; Patent No. 6037456
; GENERAL INFORMATION:
; APPLICANT: GARGER, STEPHEN
; APPLICANT: HOLTZ, R. BARRY
; APPLICANT: MCCULLOCH, MICHAEL
; APPLICANT: TURPEN, THOMAS
; TITLE OF INVENTION: A PROCESS FOR ISOLATING AND
; TITLE OF INVENTION: PURIFYING VIRUSES SOLUBLE PROTEINS AND PEPTIDES
; TITLE OF INVENTION: FROM PLANT SOURCES
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howrey & Simon
; STREET: 1299 Pennsylvania Avenue N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
```

```

: OPERATING SYSTEM: DOS
: SOFTWARE: FastSEQ for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/037.751
: FILING DATE: 10-MAR-1998
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Halluin, Albert P
: REGISTRATION NUMBER: 25,277
: REFERENCE/DOCKET NUMBER: 00801.0140.999
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 650-463-8109
: TELEFAX: 650-463-8400
: TELEX:
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 6475 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: unknown
: MOLECULE TYPE: Genomic RNA
: US-09-037-751-4

alignment_scores:
  Quality: 278.00      Length: 1895
  Ratio: 0.401        Gaps: 60
Percent Similarity: 36.570 Percent Identity: 16.570

alignment_block:
US-09-301-906-15 x US-09-037-751-4 ..
Align seg 1/1 to: US-09-037-751-4 from: 1 to: 6475

382 ThrAlaIleValAlaGlyAspArgThrArgLeuGlyGlyLeuAl 398
|||||:|||||: : : : : : : : : : : : : : : : : :
9 ACAACAAUUAACCAACAAACAAACAAACAAACAAUUAUUAU 58
: : : : : : : : : : : : : : : : : : : : : : : :
398 aserValAlaAsnAlaLeuAsnGlnGluValTyrSerValVal 415
: : : : : : : : : : : : : : : : : : : : : : : :
59 UACAAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAU 108
: : : : : : : : : : : : : : : : : : : : : : : :
415 erSerIleThrAsnArgLeuValLeuArgAspGlnSerAlaLeuSer 431
: : : : : : : : : : : : : : : : : : : : : : : :
109 ACACUGUCCGAGGAAACAAUCCUUGGUAUUAUUAUUAUUAU 158
: : : : : : : : : : : : : : : : : : : : : : : :
432 HisLeuAspThrLysLeuCysAspMetPheSerGlnArgAspAlaMetI 448
|||||:|||||: : : : : : : : : : : : : : : : : :
159 CUUUACGACACAGCGGUU...GAAGAGUUUAACGCGUCGAC..... 197
: : : : : : : : : : : : : : : : : : : : : : : :
448 eArgGluLysProSerHisArgCysAspValPheLeuLysProArgGluA 465
: : : : : : : : : : : : : : : : : : : : : : : :
198 ....CCGAGCCCAAGUGAACUUUUAACAAUUAUUAUUAUUAUUAU 243
: : : : : : : : : : : : : : : : : : : : : : : :
465 rGluLysLeuArgGluLeuPheProGluLeuSerIleGlnPheSerAsp 481
: : : : : : : : : : : : : : : : : : : : : : : :
244 CGCUUAUUGUACCCGGCGGUUUAUUAUUAUUAUUAUUAUUAUUAU 293
: : : : : : : : : : : : : : : : : : : : : : : :
482 SerValArgSerSerHisProPheAlaAsnAlaMetArgSerCys...Ph 497
: : : : : : : : : : : : : : : : : : : : : : : :
294 AGCAAAUUGCGUGCAUUCGUUGGAGGUAUUAUUAUUAUUAUUAUUA 343
: : : : : : : : : : : : : : : : : : : : : : : :
497 eAsnGlyIlePheSerArgArgCysGlyAsnValCysPheAspIleG 514
: : : : : : : : : : : : : : : : : : : : : : : :
344 GGAUAUUGCAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAU 393
: : : : : : : : : : : : : : : : : : : : : : : :
514 lYcylSerPheThrTyrHisValLysAlaGlyHisValAsnCysHisVal 530
|||||:|||||: : : : : : : : : : : : : : : : : :
394 GCGGNAUUUUGCAUUCGUUGGAGGUAUUAUUAUUAUUAUUAUUAU 443

```

```

531 CysAsnProValLeuAspValLysAspValLysArgArgIleAsnGlu... 546
||| ||| |||||:|||||: : : : : : : : : : : : : :
444 UGCAUGCCCAACCGGACGUGGAGACAUCAUGCGGACGAGCCAGAA 493
: : : : : : : : : : : : : : : : : : : : : : : :
547 .....IleLeuPheLeuSer.....ThrAlaGlyGly..... 555
|||||:|||||: : : : : : : : : : : : : : : : : :
494 AGACAGUAUUGAUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 543
: : : : : : : : : : : : : : : : : : : : : : : :
556 .....AspSerTyr.....ValSerSer 561
||| ||| : : : : : : : : : : : : : : : : : :
544 UCCCAACAUUCCAAAGGAACGAUUAUUAUUAUUAUUAUUAUUAU 593
: : : : : : : : : : : : : : : : : : : : : : : :
562 AspLeuLeuThrGluAlaAlaSerLysSerValSerTyrCysSerArgG 578
|||||:|||||: : : : : : : : : : : : : : : : : :
594 GACGUGUCUGUACAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 634
: : : : : : : : : : : : : : : : : : : : : : : :
578 userGlnAsnCysAspSerArgAlaAspAlaGlyPheMetValAspVal 595
: : : : : : : : : : : : : : : : : : : : : : : :
635 GCCGAUCCAGCAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAU 684
: : : : : : : : : : : : : : : : : : : : : : : :
595 yrAspIleSerProGlnValAlaGluAlaMetAspLysLysGlyAla 611
|||||:|||||: : : : : : : : : : : : : : : : : :
685 AUGACAUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 734
: : : : : : : : : : : : : : : : : : : : : : : :
612 LeuValPheAspIleAlaLeuMetPheProValGluLeuLeuTyrGly 628
: : : : : : : : : : : : : : : : : : : : : : : :
735 CAUACGUGUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAU 784
: : : : : : : : : : : : : : : : : : : : : : : :
628 nGlyGluValTyrLeuGluLeuAspThrLeuValLysArgGluGlyA 645
: : : : : : : : : : : : : : : : : : : : : : : :
785 UUCAUACGUCUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAU 834
: : : : : : : : : : : : : : : : : : : : : : : :
645 spTyrLeuAlaTyrAsnValGlyGlnCysGlyGluMet...TyrGluHis 660
||| ||||:|||||: : : : : : : : : : : : : : : : :
835 ACAAGUUGACCUUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAU 884
: : : : : : : : : : : : : : : : : : : : : : : :
661 SerPheSerAsnValSerGlyPhePheThrPheSerTyrValArgThr 677
|||||:|||||: : : : : : : : : : : : : : : : : :
885 AGUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAU 934
: : : : : : : : : : : : : : : : : : : : : : : :
677 rSerGlyAsnValPheLysLeuGluTyrGluGlyTyrArgCysGlyTyr 694
: : : : : : : : : : : : : : : : : : : : : : : :
935 UUAUACGAGGUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 962
: : : : : : : : : : : : : : : : : : : : : : : :
694 iHisLeuThrMetCysArgAlaGlnLysSerProGlyThrGluValThr 710
: : : : : : : : : : : : : : : : : : : : : : : :
962 ..... 962
711 TyrArgSerLeuValProSerPheValGlyLysSerLeuValPheIlePr 727
: : : : : : : : : : : : : : : : : : : : : : : :
962 ..... 962
727 oValValAlaGlySerSerValSerPheLysThrIleValLeuAspSerA 744
: : : : : : : : : : : : : : : : : : : : : : : :
962 ..... 962
744 spPheValAspArgIleTyrSerTyr.....AlaLeuAsnThrIleGly 758
: : : : : : : : : : : : : : : : : : : : : : : :
963 ..UUAUACGACGAGUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 1010
: : : : : : : : : : : : : : : : : : : : : : : :
759 ThrPheGluAsnArgThrPheGluTyrAlaValAlaGlyAlaValArgSerG 775
|||||:|||||: : : : : : : : : : : : : : : : : :
1011 ACUUUU..... 1016
: : : : : : : : : : : : : : : : : : : : : : : :
775 nLysThrHisValIleThrGlySerArgValValHisSerLysValAspI 792
: : : : : : : : : : : : : : : : : : : : : : : :
1017 .....CUUUUUAACAAAGGUGGCGCCAU..... 1040
: : : : : : : : : : : : : : : : : : : : : : : :
792 leSerProAspAspMetTrpGlyLeuValValAlaValMetAlaGlnAla 808
: : : : : : : : : : : : : : : : : : : : : : : :
1040 ..... 1040
809 IleLysAspArgAlaLysSerIleArgSerTyrAsnPheIleLysAlaSe 825

```

1041 |||||: ||| :||| ||| 1075
825 rGluGlySerLeuAlaGlyValPheLysLeuPhePheGlnThrValGlyA 842
1076 GGAAGACGCAUGCAU.....UACAAAAGACUUGCAA 1110
842 sPCysPheSerAsnAlaValSerValTyrAlaLysAlaMetValHisAsp 858
1111 UGUGCAACAGCGAGAGAUCCUUGAGGAUUAUCAUCAUCAUAUAC 1160
859 AsnPheAsnValLeuGluThrLeuMetSerMetProArgAlaPheIleAr 875
1161 UGUUUUCCAAAAGAGGAUUGGUCAUCGUACCA..... 1196
875 gLysValProGlySerValValThrIleCysThrSerGlyAlaSerA 892
1196 1196
892 sPArgLeuGluLeuArgGlyAlaPheAspIleSerLysGluThrPheGly 908
1197UUUUGCAUUCUUGGAGACUAGUAAG 1226
909 ArgLysLeuLysAsn.....SerArgLeuArgValPheSerArgAl 922
1227 AGCACCCCAAGGAGUCUUGUCCAAAGGAUUCUGUUAACA..... 1271
922 aIleValGluAspSerIleLysValMetLysAlaMetLysThrGluAspG 939
1272GUCUUAACCAUUCGAACUACCAAG 1299
939 lYlYsProIeuProIleThrGluAspSerValTyrAlaPheIleMetGly 955
1300 CCAAGUCUU.....ACAUACGCAAAUGUUUUGUUGUUGCAUUGC 1343
956 AsnValSerAsnValHisCysThrArgAlaGlyLeuLeuGlyGlySerly 972
1344 AUUCGAUCGAGGUA.....AUCAUUUACGGUGUGAC 1375
972 sAlaThrValValSerSerValSerLysGlyLeuValAlaArgGlyAlaA 989
1376 ACCGAGUCCGAAUGGGAUGUGACAAUUCUUUGUUAACA..... 1415
989 lAThrLysAlaPheSerGlyIleThrSerPhePheSerThrGlySerLeu 1005
1416UCCUUGUCCAUAGAG 1430
1006 PheTyrAspArgGlyLeuThrCgluAspGluArgLeuAspAlaLeuValAr 1022
1431 UUUUAC..... 1436
1022 gThrGluAsnAlaIleAsnSerProValGlyIleLeuGluThrSerArgV 1039
1437CUGCAUUAACUUGCCUUCUUAAGGAUGAC...U 1470
1039 aIAlaValSerLysValValAlaGlyThrLys.....GluPhe 1051
1471 UACUGAUUAGCAAGUUUAGUCUGGUGUCCGAAACGGUGCCAGCAUGUG 1520
1052 TrpSerGluValSerLeuAsnAspPheThrThrPheValLeuArgAsnLy 1068
1521 UGGGAUGAGAUUUCGUG..... 1538
1068 sValLeuIleGlyIlePheValAlaSerLeuGlyAlaAlaProIleAlaT 1085
1538 1538
1085 rPLysTyrArgArgGlyIleAlaAlaAsnAlaArgArgTyrAlaGlySer 1101
1538 1538
1102 SerTyrGluThrLeuSerSerLeuSerSerGlnAlaAlaGlyGlyLeuAr 1118

1538 1538
1118 gGlyLeuThrSerSerThrValSerGlyGlySerLeuValValArgArgG 1135
1538 1538
1135 lYPheSerSerAlaValThrValThrArgAlaThrValAlaLysArgGln 1151
1538 1538
1152 ValProLeuAlaLeuSerPheSerThrSerTyrAlaIleSerGlyCy 1168
1538 1538
1168 sSerMetLeuGlyIleTrpAlaHisAlaLeuProArgHisLeuMetPheP 1185
1538 1538
1185 hePheGlyLeuGlyThrLeuLeuGlyAlaArgAlaSerAlaAsnThrTrp 1201
1538 1538
1202 LysPheGlyGlyPheSerAsnTrpCysAlaValProGluValValTr 1218
1539GCGUUUGGGAACGCAUUCUCCUGGAAAGAGAGGCUUU 1579
1218 pArgGlyLysSerValSerLeuLeuProIleThrLeuGlyValS 1235
1580 GAACAGGAACUUUACAGA.....GUGGAGGCGCAGC 1611
1235 erLeuIleIleArgGlyLeuLeuAsnAspThrIleProGlnLeuAlaTyr 1251
1612 CAUUAAGAGAUCAAGG.....GUGCCUGAUCUA...UAU 1640
1252 ValProProValGluGlyArgAsnValTyrAspGluThrLeuArgTyrTy 1268
1641 GUG..... 1643
1268 rArgAspPheAspTyrAspGluGlyAlaGlyProSerGlyThrGlnHisG 1285
1644ACUUCACAGC 1653
1285 luAlaValProGlyAspAsnAspGlySerThrSerValSer 1301
1654 ACAGAUUA.....GUGACUGAG 1670
1302 TyrAspValValThrAsnValArgAspValGlyIleSerThrAsnGlyGl 1318
1671 UACAAGGCCUCUGGACACUUGCCUGCAUUAAGGAAGAUGCA 1720
1318 uValThrGlyGluGluThrHisSerProArgSerValGlnTyrThrT 1335
1721 AGAAACG.....GAAGUGAGUACAAUG 1743
1335 yrValGluGluGluValAlaProSerAlaAlaValAlaGluArgGlnGly 1351
1744 CACUUUCAGAG..... 1754
1352 AspProSerGlySerGlyThrAlaAspAlaMetAlaPheValGluSerVa 1368
1755UUUACGGUGUUAAGGAGUC 1774
1368 lLysLysGlyValAspAspValPheHisGlnSerSerGlyGluThrA 1385
1775 UGACAAAUUCGAGUUGAUUUUUUCCAGAU.....U 1809
1385 laArgGluValGluValAspGlyLysGlyLeuLeuProGluSerValVal 1401
1810 GCCAAUCUUUGAAGUUGAC.....CCAAUGACGCGACGGAAGGUUA 1853
1402 GlyGluAlaProThrGlnGluArgGly.....ArgAl 1412
1854 GUGCGGGUCAUGAGCAUAGAGCGGUGUCUCACAUUUUGAACGACC 1903

[illegible]

1712	uLysSerProGlyValProAsnGlnGlnProGlnSerGlnGlyAlaThrA	1729
2336	2336
1729	rgThrIleThrProLysSerGlyGlyLysAlaLeuSerGluGlySerGly	1745
2337	2337
1746	ArgGluValLysGlyArgSerThrTyrSerIle.....TrpCy	1758
2379	AGGAAG.....UAUAUGUGGCGCUUUGGAAUAUGA	2410
1758	sGluGlnAspTyrValArgLysCysGluTrpLeuArgAlaAspAsnProv	1775
2411	UGAGCAGGGUGUGGUGACAUUGGAGAAGA.....	2447
1775	alMetAlaLeuGluProAspTyrThrProMetThrPheGluValValLys	1791
2447	2447
1792	ThrGlyThrSerGluAspAlaValValGluTyrLeuLysTyrLeuAlaI	1808
2448	GUAGCUGCAGCUCUGACUGUGU.....UAUCCGACAU	2485
1808	eGlyIleGluArgThrTyrArgAlaLeuLeuMetAlaArgAsnIleAlav	1825
2486	GGCGAAACUCAGAACUCUGCGACAGCUCGUAACGAGGAGAACCGAUG	2535
1825	alThrThrAlaGluGlyValLeuLysValProAsnGlnValTyrGluSer	1841
2536	UCAGUAGCGAAAGGUGUUCU.....	2558
1842	LeuProGlyPheHisValTyrLysSerGlyThrAspLeuIlePheHis	1858
2558	2558
1858	rThrGlnAspGlyLeuArgValArgAspLeuProTyrValLeuIleAlaG	1875
2558	2558
1875	luLysGlyIlePheThrLysGlyLysAspValAspAlaValAlaLeu	1891
2559GUGGACGGAGUU.....	2570
1892	GlyAspAsnLeuPheValCysAspAspIleLeuValPheHisAspAlaI	1908
2570	2570
1908	eAsnLeuIleGlyAlaLeuLysValAlaArgCysGlyMetValGlyGlu	1925
2570	2570
1925	erPheLysSerPheGluTyrLysCysTyrAsnAlaProGlyGlyGly	1941
2571CCGGCGCUGUGG	2582
1942	LysThrThrThrLeuValAspGluPheValLysSerProAsnSerThrAl	1958
2583	AAA.....ACCAA	2590
1958	aThrIleThrAlaAsnValGlySerSerGluAspIleAsnMetAlaVal	1975
2591	AGAAAUUUUCCAGGGUUUUUGAUGAAGACUAUUUUUUAUCCUG	2640
1975	ysLys.....ArgAspProAsnLeuGluGlyLeu	1984
2641	GGAAAGCAGCGCGGAAAUAGCAAGACAGCUGCGAAAUCCACGGAU	2690
1985	AsnSerAlaThr.....ThrValAsnSerArgValValas	1996
2691	AUUGUGGCCAGGAGGACACGUUAAACCGUUGAUUUUCCUGAUGAA	2740
1996	nPhe.....IleValArgGlyMetTyrLysArgValLeuValAspGlu	2011


```

1085 rpLysTyrArgArgGlyIleAlaAlaAsnAlaArgArgTyrAlaGlySer 1101
1538 .....
1102 ScrtYrGluThrLeuSerSerLeuSerSerGlnAlaAlaGlyGlyLeuAr 1118
1538 .....
1118 gGlyLeuThrSerSerThrValSerGlyGlySerLeuValValArgArgG 1135
1538 .....
1135 lyPheSerSerAlaValThrValThrArgAlaThrValAlaLysArgGln 1151
1538 .....
1152 ValProLeuAlaLeuLeuSerPheSerThrSortYrAlaIleSerGlyCy 1168
1538 .....
1168 sSerMetLeuGlyIleTrpAlaHisAlaLeuProArgHisLeuMetPheP 1185
1538 .....
1185 hePheGlyLeuGlyThrLeuLeuGlyAlaArgAlaSerAlaAsnThrTrp 1201
1538 .....
1202 LysPheGlyGlyPheSerAsnTrpCysAlaValProGluValValTr 1218
1539 .....GCGUUGGGAACGCAUUUCCCGUGAAGAGAGACUCUU 1579
1218 pArgGlyLysSerValSerSerLeuLeuLeuProIleThrLeuGlyValS 1235
1580 GAACAGGAACUUAUCAGA.....GUGCCAGCGCAGC 1611
1235 erLeuIleIleArgGlyLeuLeuAsnAspThrIleProGlnLeuAlaTyr 1251
1612 CAUUAGAGAUCAAGG.....GUGCCUGAUCAU...UAU 1640
1252 ValProProValGluGlyArgAsnValTyrAspGluThrLeuArgTyrTy 1268
1641 GUG..... 1643
1268 rArgAspPheAspTyrAspGluGlyAlaGlyProSerGlyThrGlnHisG 1285
1644 .....ACUUCACG 1653
1285 luAlaValProGlyAspAspAsnAspGlySerThrSerSerValSerSer 1301
1654 ACAGAUUA.....GUGACUGAG 1670
1302 TyrAspValValThrAsnValArgAspValGlyIleSerThrAsnGlyGl 1318
1671 UACRAGGCCUCUGUGGACUAGCCUGCGCUAGACUAGGAGAAGAUUGGA 1720
1318 uValThrGlyGluGluGluThrHisSerProArgSerValGlnTyrThrT 1335
1721 AGAAACG.....GAACUGAUGUACAUG 1743
1335 yrValGluGluGluValAlaProSerAlaAlaValAlaGluArgGlnGly 1351
1744 CACUUUCAGAG..... 1754
1352 AspProSerGlySerGlyThrAlaAspAlaMetAlaPheValGluSerVa 1368
1755 .....UUAUCGGUGUUAAGGAGUC 1774
1368 lLysLysGlyValAspAspValPheHisGlnGlnSerSerGlyGluThrA 1385
1775 UGACAAAUUCGAUGUUGAUUUUCCACAGU.....U 1809
1385 laArgGluValGluValAspGlyLysGlyLeuLeuProGluSerValVal 1401

```

```

..... 1853
1810 GCCAAUCUUUGAAGUUGAC.....CCAAUGACGCGCAGCGAAGGUUAUA 1853
1402 GlyGluAlaProThrGlnGluArgGly.....ArgAl 1412
1854 GUCGGGCGCAUGAGCAAGAGAGCGGUCUCUCACAUUUUAGAACGACC 1903
1412 alaAspGlyAsnThrAlaGlnThrAlaValAsnGluGlyAspArgGluP 1429
1904 UACUGAGGCGAAUGUUGCG...CUAGCUUUACAGGAUCAAGAGAGGCUU 1950
1429 roValGlnSerSerLeuValSerSerProGlnAlaAspIleProLysVal 1445
1951 CAGAAAGGUGCAUUGUAGUACCUCAAGAGAAGUUGAAGAACCCGUCCAUG 2000
1446 ThrGlnSerGluValHisAlaGlnLysGluValLysGlnGluValProLe 1462
2001 AAGGUUGCG.....AUGCCAGAGAGAGAUUUACAUAU 2032
1462 uAlaThrValSerGlyAlaThrProIleValAspGluLysProAlaProS 1479
2033 AGCUGGUCUUGCGAGAUCAUCG..... 2057
1479 erValThrThrArgGlyValLysIleIleAspLysGlyLysAlaValAla 1495
2057 ..... 2057
1496 HisValAlaGluLysLysGlnValGlnValGluGlnProLysGlnArgSe 1512
2058 GAGUCGUCUUAUUCUAAAGCAGGAGAUAGAGUCUUUAGCAGGUUCA 2107
1512 rLeuThrIleAsnGluGlyLysAlaGlyLysGlnLeuCysMetPheArgT 1529
2108 UAUGCGCAGCGCAGAUUCGUUAUUCGUAAGCAGAGAGCUCGAUU... 2153
1529 hrCysSerCysGlyValGlnLeuAspValTyrAsnGluAlaThrIleAla 1545
2154 .....GUGUACACGCGGUGCCGAUUAUUAAGUU 2177
1546 ThrArgPheSerAsnAlaPheThrPheValAspAsnLeuLysGlyArgSe 1562
2178 CAGCAAAUGAAAAAC.....UUUAUCGGAUAGCCUGGUGAGCAUCACU 2218
1562 rAlaValPhePheSerLysLeuGlyGluGlyTyrThrTyrAsnGlyGlys 1579
2219 AUCUGCU..... 2225
1579 erHisValSerSerGlyTrpProArgAlaLeuGluAspIleLeuThrAla 1595
2226 .....GCGGUGUGCAUUCGUCUACAAGAU 2249
1596 IleLysTyrProSerValPheAspHisCysLeuValGlnLysTyrLysMe 1612
2250 CUCAAGAUAACAGCUGCUAUUGAC.....CUUGAAACCCGUCUCAAAGUU 2293
1612 tGlyGlyGlyValProPheHisAlaAspAspGluGluCysTyrProSerA 1629
2294 UGGA..... 2297
1629 spAsnProIleLeuThrValAsnLeuValGlyLysAlaAsnPheSerThr 1645
2297 ..... 2297
1646 LysCysArgLysGlyGlyLysValMetValIleAsnValAlaSerGlyAs 1662
2298 .....GUCUUGGAUGUUGUACUAGGAA 2320
1662 pTyrPheLeuMetProCysGlyPheGlnArgThrHisLeuHisSerValA 1679
2321 GUGGUUAAUCAACCA..... 2336
1679 snSerIleAspGluGlyArgIleSerLeuThrPheArgAlaThrArgArg 1695

```


2336 2336
1696 valPheGlyValGlyArgMetLeuGlnLeuAlaGlyValSerAspGI 1712
2336 2336
1712 uLysSerProGlyValProAsnGlnGlnProGlnSerGlnGlyAlaThrA 1729
2336 2336
1729 rgThrIleThrProLysSerGlyGlyLysAlaLeuSerGluGlySerGly 1745
2337 2337
1746 ArgGluValLysGlyArgSerThrTyrSerIle.....TrpCy 1758
2379 AGAAG.....UAUCAUGGGCGCUUUUGGAUAUGA 2410
1758 sGluGlnAspTyrValArgLysCysGluTyrLeuArgAlaAspAsnProv 1775
2411 UGACGAGGUGUGACAUCCGAGUAUGAUGGAGAAGA..... 2447
1775 alMetAlaLeuGluProAspTyrThrProMetThrPheGluValValLys 1791
2447 2447
1792 ThrGlyThrSerGluAspAlaValValGluTyrLeuLysTyrLeuAlaI 1808
2448 GUAGCUGUAGCUGAGUCUGUUGU.....UAUCCGACAU 2485
1808 eGlyIleGluArgThrTyrArgAlaLeuLeuMetAlaArgAsnIleAlav 1825
2486 GGCAGAACUCAGAACUCUGCGCAGACUGUCUUGGAACGGAGACCGCAUG 2535
1825 alThrThrAlaGluGlyValLeuLysValProAsnGlnValTyrGluSer 1841
2536 UCAGUACCGCAAGCUUGUUCU..... 2558
1842 LeuProGlyPheHisValTyrLysSerGlyThrAspLeuIlePheHis 1858
2558 2558
1858 rThrGlnAspGlyLeuArgValArgAspLeuProTyrValLeuIleAlaG 1875
2558 2558
1875 luLysGlyIlePheThrLysGlyLysAspValAspAlaValAlaLeu 1891
2559GUGGACGGAGUU..... 2570
1892 GlyAspAsnLeuPheValCysAspAspIleLeuValPheHisAspAlaI 1908
2570 2570
1908 eAsnLeuIleGlyAlaLeuLysValAlaArgCysGlyMetValGlyGluS 1925
2570 2570
1925 erPheLysSerPheGluTyrLysCysTyrAsnAlaProProGlyGlyGly 1941
2571CCGGCGCUGUGA 2582
1942 LysThrThrThrLeuValAspGluPheValLysSerProAsnSerThrAl 1958
2583 AAA.....ACCAA 2590
1958 aThrIleThrAlaAsnValGlySerSerGluAspIleAsnMetAlaValL 1975
2591 AGAAAUUCUUCAGGAGUUAUUUGAUGAAGAUCAUAAUUUUUAGUACCU 2640
1975 ysLys.....ArgAspProAsnLeuLeuGlyLeu 1984
2641 GGAAGCAGCGCGGAAAUCAUGAAGACAGACUGCGAAUCCUCCAGGGAU 2690

1985 AsnSerAlaThr.....ThrValAsnSerArgValValas 1996
2691 AUUGUGCCACGAAGGACACGUUAAAAACCGUUGAUUUUCAUGAUGAA 2740
1996 nPhe.....IleValArgGlyMetTyrLysArgValLeuValAspGluV 2011
2741 UUUUGGAAAAAGACACCGCUGUCAGUUAACAGAGGUUAUUCAUUGAUGAAG 2790
2011 alHisMetMetHisGlnGlyLeuLeuGlnLeuGlyValPheAlaThrGly 2027
2791 GGUUGAUGUUGCAUACUGGUUGUUAUUUUUUUGUGCGCAUGCAUUG 2840
2028 AlaSerGluGlyLeuPhePheGlyAspIleAsnGlnIleProPheIleAs 2044
2841 UGCAGAAUUGCAUUGUUUACGGAGACACACAGCAGAUUCCAUCAUCA 2890
2044 nArgGluLysValPheArgMetAspCysAlaValPheValProLysLysG 2061
2891 UAGAUUUCAGGAUUCGCGUACCCGCCCAUUUUUGCCAAAUUGGAUGUUG 2940
2061 luSerValValTyrThrSerLysSerTyrArgCysProLeuAspValCys 2077
2941 ACGAGUGGAGACACGCGAGAAUACUCUCCUUGUCCAGCGCGCAUGACA 2990
2078 TyrLeuLeuSerSerMetThrValArgGlyThrGluLysCysTyrProGI 2094
2991 CAUUAUCUGAAC.....AGGAGAUUAUGAGGG 3016
2094 uLysValValSerGlyLysAspLysProValValArgSerLeuSerLysA 2111
3017 CUUUGUGAUGAGC.....ACUUCUUGCUUUAAAAAGUCUGUUGCGCAGG 3060
2111 rgProIleGlyThrThrAspAspValAlaGluIleAsnAlaAspValTyr 2127
3061 AGAUGUGCGCGGAGCGCGGAGUACAUCUCCAUCAUCAAACCCUUGCAU 3110
2128 LeuCysMetThrGlnLeuGlnLysSerAspMetLysArgSerLeuLysGI 2144
3111 GGCAAGAUCUGACUUAUCCCAUUGCAU...AAAGAAGCUCUGCUUUC 3157
2144 yLysGlyLysGluThrProValMetThrValHisGluAlaGlnGlyLysT 2161
3158 AAGAGG...UAUUCAGAUUUCACACUGUGCAUGAAGUAGUAGUAGUAG 3204
2161 hrPheSerAspValValLeuPheArgThrLysLysAlaAspAspSerLeu 2177
3205 CAUACUCUGAUUUUACUAGUUAAGGUUAACCCUACACACACUCCAUCC 3254
2178 PheThr...LysGlnProHisIleLeuValGlyLeuSerArgHisThrAr 2193
3255 AUUGCAGGAGACGCCCAUGUUUUGUGCAUUGUUGCAUUGCAAGGCACAC 3304
2193 gSerLeuValTyrAlaAla.....LeuSerSerLysLeuA 2205
3305 UUCGCUCAAGUACUACACUGUUGUUAUGGAUCCUUUAUUAUUAUUAU 3354
2205 spAsp.....LysValGlyThrTyrIleSerAsp 2214
3355 GAGAUCAUGAGAAACUAGCUGUACUUAUUAUUAUUAUUAUUAUUAU 3389
seq_name: /cgnl_7/ptodata/1/ina/6_COMB.seq:US-09-037-751-1

seq_documentation_block:
; Sequence 1, Application US/09037751
; Patent No. 6037456
; GENERAL INFORMATION:
; APPLICANT: GARGER, STEPHEN
; APPLICANT: HOLTZ, R. BARRY
; APPLICANT: MCCULLOCH, MICHAEL
; APPLICANT: TURPEN, THOMAS
; TITLE OF INVENTION: A PROCESS FOR ISOLATING AND
; PURIFYING VIRUSES SOLUBLE PROTEINS AND PEPTIDES

1011	ACUUUU.		1016	1021	1026	1031	1036	1041	1046	1051	1056	1061	1066	1071	1076	1081	1086	1091	1096	1101	1106	1111	1116	1121	1126	1131	1136	1141	1146	1151	1156	1161	1166	1171	1176	1181	1186	1191	1196	1201	1206	1211	1216	1221	1226	1231	1236	1241	1246	1251	1256	1261	1266	1271	1276	1281	1286	1291	1296	1301	1306	1311	1316	1321	1326	1331	1336	1341	1346	1351	1356	1361	1366	1371	1376	1381	1386	1391	1396	1401	1406	1411	1416	1421	1426	1431	1436	1441	1446	1451	1456	1461	1466	1471	1476	1481	1486	1491	1496	1501	1506	1511	1516	1521	1526	1531	1536	1541	1546	1551	1556	1561	1566	1571	1576	1581	1586	1591	1596	1601	1606	1611	1616	1621	1626	1631	1636	1641	1646	1651	1656	1661	1666	1671	1676	1681	1686	1691	1696	1701	1706	1711	1716	1721	1726	1731	1736	1741	1746	1751	1756	1761	1766	1771	1776	1781	1786	1791	1796	1801	1806	1811	1816	1821	1826	1831	1836	1841	1846	1851	1856	1861	1866	1871	1876	1881	1886	1891	1896	1901	1906	1911	1916	1921	1926	1931	1936	1941	1946	1951	1956	1961	1966	1971	1976	1981	1986	1991	1996	2001	2006	2011	2016	2021	2026	2031	2036	2041	2046	2051	2056	2061	2066	2071	2076	2081	2086	2091	2096	2101	2106	2111	2116	2121	2126	2131	2136	2141	2146	2151	2156	2161	2166	2171	2176	2181	2186	2191	2196	2201	2206	2211	2216	2221	2226	2231	2236	2241	2246	2251	2256	2261	2266	2271	2276	2281	2286	2291	2296	2301	2306	2311	2316	2321	2326	2331	2336	2341	2346	2351	2356	2361	2366	2
------	---------	--	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	---

```
1368 lLysLysGlyValAspValPheHisGlnGlnSerSerGlyGluThrA 1385
      |||
      |||||
1775 UGACAAAUUCGAGUGUGAUUUUCCAGAU.....U 1809
      |||||
1385 laArgGluValGluValAspGlyLysGlyLeuLeuProGluSerValVal 1401
      |||||
1810 GCCAAUCUUUGGAGUGUAC.....CCNAUGACGGCAGCGAAGGUUAUA 1853
      |||||
1402 GlyGluAlaProThrGlnGluArgGly.....ArgAl 1412
      |||
1854 GUCGCGGUCAUGACGACGAGCGGUGUCUCACACAUUUAACGACC 1903
      |||||
1412 aAlaAspGlyAsnThrAlaGlnThrAlaValAsnGluGlyArgGluP 1429
      |||||
1904 UACUGAGCGGAAGUUGGC...CUAGCUUUACAGGAUCAAAGAGGCUU 1950
      |||
1429 roValGlnSerSerLeuValSerSerProGlnAlaAspIleProLysVal 1445
      |||||
1951 CAGAAGGUGCAUUGUGUUAUCCUACAGAGAAGUUGAAGACCGUCCAUG 2000
      |||
1446 ThrGlnSerGluValHisAlaGlnLysGluValLysGlnGluValProLe 1462
      |||
2001 ANGCGUUCG.....AUGGCCAGAGGAGAGUUAACAAU 2032
      |||
1462 uAlaThrValSerGlyAlaThrProIleValAspGluLysProAlaPro 1479
      |||||
2033 AGCUGGUCUUGCUGGAGAUCAUCCG..... 2057
      |||
1479 erValThrThrArgGlyValLysIleIleAspLysGlyLysAlaValAla 1495
      |||
2057 ..... 2057
1496 HisValAlaGluLysLysGlnValGlnValGluGlnProLysGlnArgSe 1512
      |||||
2058 GAGUGUCCUUAUUCUAGACGAGGAGAGUAGUCUUAAGCAGCUUA 2107
      |||||
1512 rLeuThrIleAsnGlnGlyLysAlaGlyLysGlnLeuCysMetPheArgT 1529
      |||||
2108 UAUGCGCAGCGAGAUUGUUAUUCGUAAGCAGAGUAGCUGCAU... 2153
      |||||
1529 hrCysSerCysGlyValGlnLeuAspValTyrAsnGluAlaThrIleAla 1545
      |||||
2154 .....GUGUACAGCGGUGCCGUAUAAAGUU 2177
      |||||
1546 ThrArgPheSerAsnAlaPheThrPheValAspAsnLeuLysGlyArgSe 1562
      |||||
2178 CAGCAAUUGAAAAAC.....UUUAUCGAUAGCCUGGAGCAUCACU 2218
      |||||
1562 rAlaValPhePheSerLysLeuGlyGluGlyTyrThrTyrAsnGlyLys 1579
      |||||
2219 AUCUGCU..... 2225
      |||||
1579 erHisValSerSerGlyTrpProArgAlaLeuGluAspIleLeuThrAla 1595
      |||||
2226 .....GCGGUGUCGAAUUCUGCUCAAGAU 2249
      |||||
1596 lIleLysTyrProSerValPheAspHisCysLeuValGlnLysTyrLysMe 1612
      |||||
2250 CUCAAGAUAACAGCUGCUUAUUGAC.....CUUGAAACCCGCUCAAAAGUU 2293
      |||||
1612 tGlyGlyGlyValProPheHisAlaAspGluGluCysTyrProSerA 1629
      |||
2294 UGGA..... 2297
      |||
1629 spAsnProIleLeuThrValAsnLeuValGlyLysAlaAsnPheSerThr 1645
      |||
2297 ..... 2297
      |||
1646 LysCysArgLysGlyLysValMetValIleAsnValAlaSerGlyAs 1662
      |||||
2298 .....GUCUUGGAUUGCAUCUAGGA 2320
      |||||
```

```
1662 pTyrPheLeuMetProCysGlyPheGlnArgThrHisLeuHisSerValA 1679
      |||
      |||||
2321 GUGGUUAUCAAACCA..... 2336
      |||
1679 snSerIleAspGluGlyArgIleSerLeuThrPheArgAlaThrArgArg 1695
      |||
2336 ..... 2336
      |||
1696 valPheGlyValGlyArgMetLeuGlnLeuAlaGlyGlyValSerAspG1 1712
      |||
2336 ..... 2336
      |||
1712 uLysSerProGlyValProAsnGlnGlnProGlnSerGlnGlyAlaThrA 1729
      |||
2336 ..... 2336
      |||
1729 rgThrIleThrProLysSerGlyLysAlaLeuSerGluGlySerGly 1745
      |||||
2337 .....ACGCCAAGAGUCAGUAGGUGUGUUGAANAACCCACGCG 2378
      |||||
1746 ArgGluValLysGlyArgSerThrTyrSerIle.....TyrCy 1758
      |||||
2379 AGGAAG.....UAUCAUGUGCGCUUUUGGAUAUGA 2410
      |||
1758 sGluGlnAspTyrValArgLysCysGluTrpLeuArgAlaAspAsnProV 1775
      |||||
2411 UGAGCAGGGUGUGGAGACAUUGCAUGAUUGGAGAAGA..... 2447
      |||
1775 aIMetAlaLeuGluProAspTyrThrProMetThrPheGluValLys 1791
      |||
2447 ..... 2447
      |||
1792 ThrGlyThrSerGluAspAlaValValGluTyrLeuLysTyrLeuAlaI1 1808
      |||||
2448 GUAGCUGUUAUCUCUGAGUCUGUU.....UAUCCCAU 2485
      |||
1808 eGlyIleGluArgThrTyrArgAlaLeuMetAlaArgAsnIleAlav 1825
      |||||
2486 GCGAAACUCAGAACUCUGCGCAGACUGUCUUGAAACGAGAACCGCAUG 2535
      |||||
1825 aIThrThrAlaGluGlyValLeuLysValProAsnGlnValTyrGluSer 1841
      |||||
2536 UCAGUAGCGCAAGGUUGUUCUU..... 2558
      |||
1842 LeuProGlyPheHisValTyrLysSerGlyThrAspLeuIlePheHisSe 1858
      |||
2558 ..... 2558
      |||
1858 rThrGlnAspGlyLeuArgValArgAspLeuProTyrValLeuIleAlaG 1875
      |||
2558 ..... 2558
      |||
1875 lLysGlyIlePheThrLysGlyLysAspValAspAlaValAlaLeu 1891
      |||||
2559 .....GUGGACGGAGUU..... 2570
      |||
1892 GlyAspAsnLeuPheValCysAspIleLeuValPheHisAspAlaI1 1908
      |||
2570 ..... 2570
      |||
1908 eAsnLeuIleGlyAlaLeuLysValAlaArgCysGlyMetValGlyGlu 1925
      |||
2570 ..... 2570
      |||
1925 erPheLysSerPheGluTyrLysCysTyrAsnAlaProGlyGlyGly 1941
      |||||
2571 .....CCGGGUGUGGA 2582
      |||
1942 LysThrThrThrLeuValAspGluPheValLysSerProAsnSerThrAl 1958
      |||
2583 AAA.....ACCAA 2590
      |||
1958 aThrIleThrAlaAsnValGlySerSerGluAspIleAsnMetAlaValL 1975
      |||
```



```

198 ....CGCAGGCCCAAGGUGAACUUUUUCAAAAAGUAUAAGCGGAGGACGACA 243
465 rgGluLysLeuArgGluLeuPheProGluLeuSerIleGlnPheSerAsp 481
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
244 CCUUUUUUGCUACCCGGCGGUACUCCAGAAUUCCAAUUUAUUUAUAAC 293
482 SerValArgSerSerHisProPheAlaAsnAlaMetArgSerCys...Ph 497
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
294 ACCCAAAUUGCCGUGCAUUCGUGCAGGUGGAUUGCGCAUUCUUAGAACU 343
497 eAsnGlyIlePheSerArgArgCysGlyAsnValCysPhePheAspIleG 514
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
344 GGAUAUCUGAUGAUGCAAAUUCUCCAGCGGAUCAUGGAUUAUUAUUAUAG 393
514 lyGlySerPheThrTyrHisValLysAlaGlyHisValAsnCysHisVal 530
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
394 CGGGCAUUUUGCAUCGCAUCUGUUCAAGGGACGAGCAUUAUUAUUAUUA 443
531 CysAsnProValLeuAspValLysAspValLysArgArgIleAsnGlu... 546
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
444 UGCAUGCCCAACCGGAGGUGUUGGAGACAUCAUGCGGCACGAGGCCAGAA 493
547 .....IleLeuPheLeuSer.....ThrAlaGlyGly..... 555
494 AGACAGUAUUGAACUAUACCUUUUAGCGUACGAGAGGGGGGAAACAG 543
556 .....AspSerTyr.....ValSerSer 561
544 UCCCCAACUUCCAAAGGAACCAUUUGACAGAUACGCAGAAAUUCCUGAA 593
562 AspLeuLeuThrGluAlaAlaSerLysSerValSerTyrCysSerArgG1 578
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
594 CACGCGUCUGUCACAAUUAUCCAGACAAUUGCGACAU.....CA 634
578 userGlnAsnCysAspSerArgAlaAspAlaGlyPheMetValAspValT 595
635 GCCGAGGAGCAUAUCAGGAGUGUUGCCAUUGCGGCUACACAGCAUAU 684
595 yrAspIleSerProGlnValAlaGluAlaMetAspLysLysGlyAla 611
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
685 AUGACAUACCCAGCCGAGUUGCGGGCGCACUCUUGAGAAAUUUGUC 734
612 LeuValPheAspIleAlaLeuMetPheProValGluLeuLeuTyrGlyAs 628
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
735 CAUACGUGCUAUGCGCGUUUCCACUUCUCCGAGAACCUUGCUUGAAGA 784
628 nGlyGluValTyrLeuGluLeuAspThrLeuValLysArgGluGlyAla 645
785 DUCAUACGUAUUUGGAGCAAAUACCGGUGUUUUUUGCGCGGAGGAG 834
645 spTyrLeuAlaTyrAsnValGlyGlnCysGlyGluMet...TyrGluHis 660
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
835 ACAAGUACCCUUUUUUUUGCAAGAGAGAGUACUCUUAUUUAUUGUCAU 884
661 SerPheSerAsnValSerGlyPhePheThrPheSerTyrValArgThrSe 677
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
885 AGUUAUUAUUUAUUUAUUGUGUGCAAAACUUAUCCCGCGCUC 934
677 rSerGlyAsnValPheLysLeuGluTyrGluGlyTyrArgCysGlyTyrH 694
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
935 UAAUAGAGGUGGUUAUAGAGGAGUUU..... 962
694 sHisLeuThrMetCysArgAlaGlnIlySerProGlyThrGluValThr 710
962 ..... 962
711 TyrArgSerLeuValProSerPheValGlyLysSerLeuValPheIlePr 727
962 ..... 962
727 oValValAlaGlySerSerValSerPheLysThrIleValLeuAspSera 744
962 ..... 962

```

```

744 spPheValAspArgIleTyrSerTyr.....AlaLeuAsnThrIleGly 758
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
963 ..UUAUGACACCAGAGUAAUACUGGUUUUGUAAGUUUUCUAGAAUAGAU 1010
759 ThrPheGluAsnArgThrPheGluTyrAlaValGlyAlaValArgSerG1 775
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1011 ACUUUU..... 1016
775 nLysThrHisValIleThrGlySerArgValValHisSerLysValAspI 792
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1017 .....CUUUUGUACAAAGGUGGCGCCAU..... 1040
792 leSerProAspAspMetTrpGlyLeuValValAlaValMetAlaGlnAla 808
1040 ..... 1040
809 IleLysAspArgAlaLysSerIleArgSerTyrAsnPheIleLysAlaSe 825
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1041 .....AAAAGUGAGAUAGUGAGCAGUUUUUAUACUGCAAU 1075
825 rGluGlySerLeuAlaGlyValPheLysLeuPhePheGlnThrValGlyA 842
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1076 GGAAGACGCAUGGCAU.....UACAAAAGACUCUUGCAA 1110
842 spCysPheSerAsnAlaValSerValTyrAlaLysAlaMetValHisasp 858
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1111 UGUGCAACAGCAGAGAAUCCUUGAGGAUUAUCAUCAGCAUUAUAC 1160
859 AsnPheAsnValLeuGluThrLeuMetSerMetProArgAlaPheIleAr 875
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1161 UGUUUUCCAAAAGAGGGAUAGUGUCAUGUACCA..... 1196
875 gLysValProGlySerValValThrIleCysThrSerGlyAlaSerA 892
1196 ..... 1196
892 spArgLeuGluLeuArgGlyAlaPheAspIleSerLysGluThrPheGly 908
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1197 .....UUUUUGCAUUAUUUGGAGACUAGUAAAG 1226
909 ArgLysLysLeuLysAsn.....SerArgLeuArgValPheSerArgAl 922
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1227 AGGAGCGCAGAGAGUUCUAGUGUCCAAAGGAUUCUGUUUUAACA..... 1271
922 aIleValGluAspSerIleLysValMetLysAlaMetLysThrGluAspG 939
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1272 .....GUGCUUAACCAUUCGAAACAUUACCCAGG 1299
939 lyLysProLeuProIleThrGluAspSerValTyrAlaPheIleMetGly 955
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1300 CGAAAGCUCUU.....ACAUACGCAAAUUGUUUUGUC 1331
956 AsnValSerAsnValHisCysThrArgAlaGlyLeuLeuGlyGlySerLy 972
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1332 CUUGCGAAUUCGUACGA...UCGAGGUA...AUCAUUUACGGUGUGAC 1375
972 sAlaThrValValSerSerValSerLysGlyLeuValAlaArgGlyAlaA 989
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1376 AGCAGGUGCCGAAUUGGAGUGGACAAUUCUUUUUUAACA..... 1415
989 laThrLysAlaPheSerGlyIleThrSerPhePheSerThrGlySerLeu 1005
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1416 .....UCCUUGUCCAGACG 1430
1006 PheTyrAspArgGlyLeuThrGluAspGluArgLeuAspAlaLeuValAr 1022
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1431 UUUUAC..... 1436
1022 gThrGluAsnAlaIleAsnSerProValGlyIleLeuGluThrSerArgV 1039
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1437 .....CUGCAUCUACUAGCUUGCCGUUUAAGGAUGAC...U 1470

```

```

1039 aAlaValSerLysValAlaGlyThrLys.....GluPhe 1051
1471 UACUGAUUACAAUUUAGUCUGGUGGAAACGGUGGCCAGCAGUG 1520
1052 TrpSerGluValSerLeuAsnAspPheThrThrPheValLeuArgAsnLy 1068
1521 UGGAGUAGAGAUUUGCUG..... 1538
1068 sValLeuIleGlyPhePheValAlaSerLeuGlyAlaAlaProIleAlaT 1085
1538 ..... 1538
1085 rPlyTyrArgArgGlyIleAlaAlaAsnAlaArgArgTyrAlaGlySer 1101
1538 ..... 1538
1102 SerTyrGluThrLeuSerSerLeuSerSerGlnAlaAlaGlyGlyLeuAr 1118
1538 ..... 1538
1118 gGlyLeuThrSerSerThrValSerGlySerLeuValValArgArg 1135
1538 ..... 1538
1135 lYPheSerSerAlaValThrValThrArgAlaThrValAlaLysArgGln 1151
1538 ..... 1538
1152 ValProLeuAlaLeuLeuSerPheSerThrSerTyrAlaIleSerGlyCy 1168
1538 ..... 1538
1168 sSerMetLeuGlyIleTrpAlaHisAlaLeuProArgHisLeuMetPheP 1185
1538 ..... 1538
1185 hePheGlyLeuGlyThrLeuLeuGlyAlaArgAlaSerAlaAsnThrTrp 1201
1538 ..... 1538
1202 LysPheGlyGlyPheSerAsnAsnTrpCysAlaValProGluValValTr 1218
1539 .....GCUUUUGGAACGCAUUUCCUCCGUGAAGAGAGACUCUU 1579
1218 pArgGlyLysSerValSerSerLeuLeuLeuProIleThrLeuGlyValS 1235
1580 GAACAGGAACUUUCAGA.....GUGGCAGGCGACG 1611
1235 erLeuIleIleArgGlyLeuLeuAsnAspThrIleProGlnLeuAlaTyr 1251
1612 CAUUAGAGAUUCAGG.....GUGCCUGAUCUA...UAU 1640
1252 ValProValGluGlyArgAsnValTyrAspGluThrLeuArgTyrTy 1268
1641 GUG..... 1643
1268 rArgAspPheAspTyrAspGluGlyAlaGlyProSerGlyThrGlnHisg 1285
1644 .....ACCUCCACG 1653
1285 luAlaValProGlyAspAspAsnAspGlySerThrSerValSerSer 1301
1654 ACAGAUUA.....GUGACUGAG 1670
1302 TyrAspValValThrAsnValArgAspValGlyIleSerThrAsnGlyGl 1318
1671 UACAAGCCUCUGUGGACAUCCUGCGUUGACAUUAGGAAGAAGGAUGGA 1720
1318 uValThrGlyGluGluThrHisSerProArgSerValGlnTyrThrT 1335
1721 AGAAGC.....GAAGUGAGUACAUAUG 1743
1335 yrValGluGluGluValAlaProSerAlaAlaValAlaGluArgGlnGly 1351

```

```

1744 CACUUUCAGAG..... 1754
1352 AspProSerGlySerGlyThrAlaAspAlaMetAlaPheValGluSerVa 1368
1755 .....UUAUCGGUGUUUAGGGAGUC 1774
1368 lLysLysGlyValAspAspValPheHisGlnGlnSerSerGlyGluThrA 1385
1775 UGACAAAUUCGUGUUGAUUUUUUCCCAUG.....U 1809
1385 laArgGluValGluValAspGlyLysGlyLeuLeuProGluSerValVal 1401
1810 GCCAAUCUUUGGAUGUAC.....CCAAUGCCGAGGAGGAGGUUAUA 1853
1402 GlyGluAlaProThrGlnGluArgGly.....ArgAl 1412
1854 GUCGCGGUCAUGAGCAAGAGAGCGGUCUCACAUUUUGAACGACC 1903
1412 aAlaAspGlyAsnThrAlaGlnThrAlaValAsnGluGlyAspArgGluP 1429
1904 UACUGAGGCGGAUUGUGCG...CUAGCUUACAGGAUCAAGAGAAGGCUU 1950
1429 roValGlnSerSerLeuValSerSerProGlnAlaAspIleProLysVal 1445
1951 CAGAAGUGCAUUGGUUACCUCAAGAGAAGUUGAAGAACCCGUCCAUG 2000
1446 ThrGlnSerGluValHisAlaGlnLysGluValLysGlnGluValProLe 1462
2001 AAGGUGUUG.....AUGCCAGAGGAGAGAUUACAUAU 2032
1462 uAlaThrValSerGlyAlaThrProIleValAspGluLysProAlaProS 1479
2033 AGCUGGUCUUGCGAGAUCAUCG..... 2057
1479 erValThrThrArgGlyValLysIleIleAspLysGlyLysAlaValAla 1495
2057 ..... 2057
1496 HisValAlaGluLysLysGlnValGlnValGlnProLysGlnArgse 1512
2058 GAGUGUCUUAUUCUAGAAGACGAGGAGAUAGUCUUUAGAGAGAUUUA 2107
1512 rLeuThrIleAsnGluGlyLysAlaGlyLysGlnLeuCysMetPheArgT 1529
2108 UAUGGCGACGCGCAGAUUCGUUAUUCGAGACAGAGAGCUCGAU... 2153
1529 hrCysSerCysGlyValGlnLeuAspValTyrAsnGluAlaThrIleAla 1545
2154 .....GUGUACACGGGUGCCGAUUAAGUU 2177
1546 ThrArgPheSerAsnAlaPheThrPheValAspAsnLeuLysGlyArgse 1562
2178 CAGCAAAUGAAGAAAC.....UUUAUGCAUAGCCUGGAGCAUCACU 2218
1562 rAlaValPhePheSerLysLeuGlyGluGlyTyrThrTyrAsnGlyGlys 1579
2219 AUCUGCU..... 2225
1579 erHisValSerSerGlyTrpProArgAlaLeuGluAspIleLeuThrAla 1595
2226 .....CGGUGUGGAAUUCUGUCAAGAU 2249
1596 IleLysTyrProSerValPheAspHisCysLeuValGlnLysTyrLysme 1612
2250 CUCAAGAAUACAGCUGCUAUUGAC.....CUUGAAACCCGCUCAAAAGUU 2293
1612 tGlyGlyGlyValProPheHisAlaAspAspGluGluCysTyrProSerA 1629
2294 UGGA..... 2297
1629 sPasnProIleLeuThrValAsnLeuValGlyLysAlaAsnPheSerThr 1645

```

[illegible][illegible]


```
962 ..... 962
727 oValValAlaGlySerSerValSerPheLysThrIleValLeuAspSerA 744
962 ..... 962
744 spPheValAspArgIleTyrSerTyr.....AlaLeuAsnThrIleGly 758
:::||||| ::::: ::::: |||
963 ..UUAUCACACAGAGUAUAUACUGGUUUUGUAAGUUUUUUAAGAAUAGAU 1010
759 ThrPheGluAsnArgThrPheGluTyrAlaValGlyAlaValArgSerGI 775
|||||
1011 ACUUUU..... 1016
775 nLysThrHisValIleThrGlySerArgValValHisSerLysValAspI 792
::: ::::: |||:::||||
1017 .....CUUUUGUACAAGGUGUGCCCAU..... 1040
792 leSerProAspAspMetTrpGlyLeuValValAlaValMetAlaGlnAla 808
1040 ..... 1040
809 IleLysAspArgAlaLysSerIleArgSerTyrAsnPheIleLysAlaSe 825
|||||::: ||| ::::: |||
1041 .....AAAAGUGAUAAGAGAGCAAGUUUUUAUCUGCAU 1075
825 rGluGlySerLeuAlaGlyValPheLysLeuPheGlnThrValGlyA 842
||| ::: ::::: |||:::||||
1076 GGAAGACGCAUGGCANU.....UACAAAAGACUCUUGCAA 1110
842 spCysPheSerAsnAlaValSerValTyrAlaLysAlaMetValHisAsp 858
||| ||::: ::: ::::: |||:::
1111 UGUGCAACAGCAGAGAGAAUCCUUGAGGAUUAUCAUCAUGCAUUAUAC 1160
859 AsnPheAsnValLeuGluThrLeuMetSerMetProArgAlaPheIleAr 875
||| ::::: ::::: |||
1161 UGGUUUCCCAAAAGAGGGAUGGUCAUGCA..... 1196
875 gLysValProGlySerValValIleCysThrSerGlyAlaSerA 892
1196 ..... 1196
892 spArgLeuGluLeuArgGlyAlaPheAspIleSerLysGluThrPheGly 908
||||| ||| ||| |||
1197 .....UUAUUGCAUUAUUGGAGACUAGUAAG 1226
909 ArgLysLeuLysAsn.....SerArgLeuArgValPheSerArgAl 922
||| ||::: |||::: |||::: |||:::
1227 AGGACGCGCAAGGAGUCUUAAGUGUCCAAAGGAUUUCUGUUUACA..... 1271
922 aIleValGluAspSerIleLysValMetLysAlaMetLysThrGluAspG 939
1272 .....GUGCUUACCAUAUUGCAUACCAUCCAGG 1299
939 lyLysProLeuProIleThrGluAspSerValTyrAlaPheIleMetGly 955
::: ||| ::::: |||::: |||::: |||:::
1300 CGAAAGCUCUU.....ACAUACGCAAAUUGUUUGUCC 1331
956 AsnValSerAsnValHisCysThrArgAlaGlyLeuLeuGlySerLy 972
|||||::: ::: |||::: |||::: |||
1332 CUUGUCGAUUGAUUCA.....UCGAGGUA...AUCAUUAACGGUGUGAC 1375
972 sAlaThrValValSerSerValLysGlyLeuValAlaArgGlyAlaA 989
|||
1376 AGCAGGUCCAAUGGGAGUGGACAAAUUUUUUACAA..... 1415
989 laThrLysAlaPheSerGlyIleThrSerPhePheSerThrGlySerLeu 1005
||| :::
1416 .....UCCUUGUCCAUGACG 1430
1006 PheTyrAspArgGlyLeuThrGluAspGluArgLeuAspAlaLeuValAr 1022
|||||
```

1318 uValThrGlyGluGluThrHisSerProArgSerValGlnTyrThrT 1335
| |||
1721 AGAAGCGAAGUGAUGAACAU 1743
1335 yRValGluGluGluValAlaProSerAlaValAlaGluArgGlnGly 1351
: : : : :
1744 CACUUCACAG 1754
1352 AspProSerGlySerGlyThrAlaAspAlaMetAlaPheValGluSerVa 1368
: : : : :
1755UUAUCGGUGUUAAGGAGUC 1774
1368 lLysLysGlyValAspValPheHisGlnGlnSerSerGlyGluThra 1385
|||
1775 UGACAAAUCCGAUGUGUUGUUUCCAGAGU 1809
1385 lAArgGluValGluValAspGlyLysGlyLeuLeuProGluSerValVal 1401
: : : : :
1810 GCCAAUCUUGGAAGUUGACCCAAUGACGCCAGCAAGGUUAUA 1853
1402 GlyGluAlaProThrGlnGluArgGlyArgAl 1412
: : : : :
1854 GUCGCGUCAGACGAUGAGAGCGGUCGUCACUCACAUUUGAACGACC 1903
1412 aLaAspGlyAsnThrAlaGlnThrAlaValAsnGluGlyAspArgGluP 1429
: : : : :
1904 UACUGAGGCGNAUGUUGCGCUACGUUACAGGAUCAAGAGGCGUU 1950
1429 roValGlnSerSerLeuValSerProGlnAlaAspIleProLysVal 1445
: : : : :
1951 CAGAAGGCUAUGGUGUAGUUUCUAGAAGAGUUGAACCGUCCAU 2000
1446 ThrGlnSerGluValHisAlaGlnLysGluValLysGlnGluValProLe 1462
|||
2001 AAGGGUUGCAUGGCCAGAGGAGUUAACAUAU 2032
1462 uAlaThrValSerGlyAlaThrProIleValAspGluLysProAlaProS 1479
||| : : : : :
2033 AGCUGUCUGUGGAGAUCAUCCG 2057
1479 erValThrThrArgGlyValLysIleAspLysGlyLysAlaValAla 1495
2057 2057
1496 HisValAlaGluLysLysGlnValGlnValGluGlnProLysGlnArgSe 1512
: : : : :
2058 GAGUCGUCCUAUUCUAAAGAACGAGGAGUAGAGUCUUUAGAGCAGUUUCA 2107
1512 rLeuThrIleAsnGluGlyLysAlaGlyLysGlnLeuCysMetPheArgT 1529
: : : : :
2108 UAUGGCGACGGCAGAUUCGUUAAUUCGUAAGCAGAGCAGUCGUAU 2153
1529 hrCysSerCysGlyValGlnLeuAspValTyrAsnGluAlaThrIleAla 1545
|||
2154GUGUACACGGGUGCCGUAUUAAGUU 2177
1546 ThrArgPheSerAsnAlaPheThrPheValAspAsnLeuLysGlyArgSe 1562
: : : : :
2178 CAGCAAAUGAAAAACUUUAUCGAUAGCCUGUAGCAUCACU 2218
1562 rAlaValPhePheSerLysLeuGlyGluGlyTyrThrTyrAsnGlyGlys 1579
: : : : :
2219 AUCUGCU 2225
1579 erHisValSerSerGlyTyrProArgAlaLeuGluAspIleLeuThrAla 1595
||| : : : : :
2226GCGGUGUCGAAUCUCUCCAGAU 2249
1596 IleLysTyrProSerValPheAspHisCysLeuValGlnLysTyrLysMe 1612
: : : : :
2250 CUCAAGAAUACAGCCUUAUGACCUUGAAACCCGCUCAAGUU 2293

1612 tGlyGlyGlyValProPheHisAlaAspaspGluGluCysTyrProSera 1629
: : : : :
2294 UGGA 2297
1629 sPAsnProIleLeuThrValAsnLeuValGlyLysAlaAsnPheSerThr 1645
: : : : :
2297 2297
1646 LysCysArgLysGlyGlyLysValMetValIleAsnValAlaSerGlyAs 1662
||| : : : : :
2298GUCUUGAUGUUGCAUCUAGGAA 2320
1662 pTyrPheLeuMetProCysGlyPheGlnArgThrHisLeuHisSerVala 1679
: : : : :
2321 GUGGUUAUCAAACCA 2336
1679 snSerIleAspGluGlyArgIleSerLeuThrPheArgAlaThrArgArg 1695
2336 2336
1696 ValPheGlyValGlyArgMetLeuGlnLeuAlaGlyValSerAspG1 1712
2336 2336
1712 uLysSerProGlyValProAsnGlnProGlnSerGlnGlyAlaThra 1729
2336 2336
1729 rgThrIleThrProLysSerGlyLysAlaLeuSerGluGlySerGly 1745
||| : : : : :
2337ACGGCCAAAGACUAGCGUGUGUUGUAAACCCACGCG 2378
1746 ArgGluValLysGlyArgSerThrTyrSerIleTrpCy 1758
||| : : : : :
2379 AGGAAGUAUCAUGUGCGCUUUGGAUAUGA 2410
1758 sGluGlnAspTyrValArgLysCysGluThrPleuArgAlaAspAsnProv 1775
||| : : : : :
2411 UGAGCAGGGUGUGGACACUAGCGAUGAUUGGAGAAGA 2447
1775 aMetAlaLeuGluProAspTyrThrProMetThrPheGluValValLys 1791
2447 2447
1792 ThrGlyThrSerGluAspAlaValValGluTyrLeuLysTyrLeuAlaI1 1808
: : : : :
2448 GUAGCUGUUGACUCUGAGUCUGUUGUUAUCCGACAU 2485
1808 eGlyIleGluArgThrTyrArgAlaLeuLeuMetAlaArgAsnIleAlav 1825
: : : : :
2486 GCGAAACUCAGAACUCUGCGACAGCUGCUUCAAACCGGAGACCCGAUG 2535
1825 alThrThrAlaGluGlyValLeuLysValProAsnGlnValTyrGluSer 1841
||| : : : : :
2536 UCAGUAGCCGCAAGGUGUUCUU 2558
1842 LeuProGlyPheHisValTyrLysSerGlyThrAspLeuIlePheHisSe 1858
2558 2558
1858 rThrGlnAspGlyLeuArgValArgAspLeuProTyrValLeuIleAlaG 1875
2558 2558
1875 luLysGlyIlePheThrLysGlyLysAspValAspAlaValValAlaLeu 1891
||| : : : : :
2559GUGGACGGAGUU 2570
1892 GlyAspAsnLeuPheValCysAspIleLeuValPheHisAspAlaI1 1908
2570 2570
1908 eAsnLeuIleGlyAlaLeuLysValAlaArgCysGlyMetValGlyGluS 1925

[illegible][illegible]

```

1083 eAlaTrpLysTyrArgArgGlyIleAlaAlaAsnAlaArgTyrAlaG 1100
1507 .....GACCAGGCCATGACAACAGCGCCTATGAGG 1537
1100 lYserSerTyrGluThrLeuSerSerLeuSerSerGlnAlaGlyGly 1116
1538 GGTCTGATCTTACCCTGCTGATCGCTATTAGTGACATATCTGGG... 1584
1117 LeuArgGlyLeuThrSerSerThrValSerGlyGlySerLeu... 1130
1585 .....TCCTACGTAGTCCCTGGACATGCTCCCAACCGGT 1619
1131 .....ValValArgArgGlyPheSerS 1138
1620 TTACCAAGCCCTTGACCTCCCGCTGAGATTGTGGCTCGTCGAGCGCGC 1669
1138 eAlaValThrValThrArgAlaThrValAlaLysArg.....Gln 1151
1670 TCACCCCCACAGTAAAGGTCCTCCAGGTCGACGGCGGATCGATTGTGAG 1719
1152 ValProLeuAlaLeuSerPheSerThrSerTyrAlaIleSerGlyCy 1168
1720 ACCCTTCTCGGTAAATAAACCTTCCCGCATGCTTT...GTTGACGGGC 1766
1168 sSerMetLeuGlyIleTrpAlaHisAlaLeuProArgHisLeuMetPheP 1185
1767 GGT.....TTAGAGACTAATGGCCAGAGCGCCACAATCTCTCT 1807
1185 hePheGlyLeuGlyThrLeuLeuGlyAlaArgAlaSerAlaAsnThrTrp 1201
1808 TT.....GATCCAGTCAGAGCCATG 1830
1202 LysPheGlyGlyPheSerAsnAsnTrpCysAlaValProGluValValTr 1218
1831 GCGCGCGCCCTTTCAGT..... 1848
1218 pArgGlyLysSerValSerSerLeuLeuLeuProIleThrLeuGlyValS 1235
1848 ..... 1848
1235 eLeuIleIleArgGlyLeuLeuAsnAspThrIleProGlnLeuAlaTyr 1251
1849 .....CTCACCTAT 1857
1252 ValProValGluGlyArgAsnValTyrAspGluThrLeuArgTyrTy 1268
1858 GCGGCTCTGCTGCTGGGCTGGAGGTG.....CGCTATGT 1892
1268 rArgAspPheAspTyrAspGluGlyAlaGlyProSerGlyThrGlnHisG 1285
1893 C.....GCCGCGGGCTTGACCAC 1912
1285 lAlaVal.....ProGlyAspAspAsnAspGlySerThrSerServal 1299
1913 GGGCGGTTTTTCCCGCCCGC..... 1932
1300 SerSerTyrAspValValThrAsnValArgAspValGlyIleSerThrAs 1316
1932 ..... 1932
1316 nGlyGluValThrGlyGluGluThrHisSerProArgServalGlnTr 1333
1933 .....GTTTCACCCCGGTCA..... 1947
1333 yThrTyrValGluGluGluValAlaProSerSerAlaValAlaGluArg 1349
1948 .....GCCCCCT..... 1953
1350 GlnGlyAspProSerGlySerGlyThrAlaAspAlaMetAlaPheValGl 1366
1954 .....GCCGAGTCACCGCCTTCTGTTC 1976
1366 uSerValLysLysGlyValAspValPheHisGlnGlnSerSerGlyG 1383
1977 TGCCTATACAGG.....TTTAATCGCAG..... 2001
1383 luThrAlaArgGluValGluValAspGlyLys...GlyLeuLeuProGlu 1398
2002 ....GCCAGCGCCTTTCGCTGACCGGTAATTTTGGTTCCATCCTGAG 2046
1399 SerValValGlyGlu...AlaProThrGlnGluArgGlyArgAlaAlaAs 1414
2047 GGCTCCTTGGCCCTTGGCCCG.....TTTTCGCC 2078
1414 pGlyAsnThrAlaGlnThrAlaValAsnGluGlyAspArgGluProVal. 1430
2079 CGGCGATGTTGGGATCGGCT.....AATCCATTCT 2110
1431 .....GlnSerSerLeuValSerSerProGlnAlaAsp.....Ile 1442
2111 GTGGCGAGAGACACTTTACACCGCACCTTGGTCGGAGTTGATGCTGT 2160
1443 ProLysValThrGlnSerGluValHisAlaGlnLysGluValLysGlnGl 1459
2161 CCTAGTCCAGCCAGCCGACTTAGGTTTACATCTGAGCCTTCTATACC 2210
1459 uValProLeuAlaThrValSerGlyAlaThrProIleValAspGluLysP 1476
2211 TAGTAGGGCGCCACACCTACCCCGCGCCCTCTACCCCGCCCTGCAC 2260
1476 roAlaProSerValThr.....ThrArgGlyValLysIleIle 1488
2261 CGGATCTTCCCTTACTCTCTGCTCCGCGCGGT..... 2298
1489 AspLysGlyLysAlaValAlaHisValAlaGluLysLysGlnValGlnVa 1505
2299 .....GAGCGGCTCTGGCGCTACCGCGCGCCAGCCCATACCCA 2342
1505 lGluGlnProLysGlnArgSerLeuThrIleAsnGluGlyLysAlaGlyL 1522
2343 CCAGACGGCGCGCATCGCGCTGCTC..... 2370
1522 ysGlnLeuCysMetPheArgThrCysSerCysGlyValGlnLeuAspVal 1538
2370 ..... 2370
1539 TyrAsnGluAlaThrIleAlaThrArgPheSerAsnAlaPheThrPheVa 1555
2370 ..... 2370
1555 lAspAsnLeuLysGlyArgSerAlaValPhePheSerLysLeuGlyGluG 1572
2370 ..... 2370
1572 lyTyrThrTyrAsnGlyGlySerHisValSerSerGlyTrpProArgAla 1588
2371 ..TTTACCTACCGGATGCTCTAAGGTGTTCGCGGC.....TCGCTG 2412
1589 LeuGluAspIleLeuThrAlaIleLysTyrProSerValPheAspHisCy 1605
2413 TTTGAGTCGACATGTACCTGGCTGCTTAACGCTCTAATGTTGACCAC... 2460
1605 sLeuValGlnLysTyrLysMetGlyGlyValProPheHisAlaAspa 1622
2461 .....CGCCCTGGCGTGGGCTC...TGTCATGCATTTT 2491
1622 spGluGluCysTyrProSerAspAsnProIleLeuThrValAsnLeuVal 1638
2492 ACCAGAGG...TACCCCGCTCC..... 2511
1639 GlyLysAlaAsnPheSerThrLysCysArgLysGlyGlyLysValMetVa 1655
2511 ..... 2511
1655 lIleAsnValAlaSerGlyAspTyrPheLeuMetProCysGlyPheGlnA 1672

```

OM of: US-09-301-906-15 to: N_Geneseq_36:* out_format : pfs
Date: Feb 18, 2001 7:30 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
-MODEL=framed-p2n.model -DEV=rlp
-Q/cgnl_1/USPTO-us09301906/runat_15022001_103219_9340/app_query.fasta_1.2332
-DB=N_Geneseq_36 -QFMT=fastap -SUFFIX=ring -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-GAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=EXT -MINLEN=0 -MAXLEN=200000000
-USER=US09301906_cgcnl_1_0 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:
Query: US-09-301-906-15
Query length: 2237
Database: N_Geneseq_36:*
Database sequences: 480022
Database length: 187831343
Search time (sec): 212.710000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	! Documentation
/cgnl_8/gcgdata/geneseq/NA2000.DAT:249208 +			11520.00	16457.74	0.0
/cgnl_8/gcgdata/geneseq/NA2000.DAT:249200 +			11520.00	16446.22	0.0
/cgnl_8/gcgdata/geneseq/NA1997.DAT:T72214 +			5247.00	7478.02	0.0
/cgnl_8/gcgdata/geneseq/NA2000.DAT:249205 +			1495.00	2122.11	2.1e-110
/cgnl_8/gcgdata/geneseq/NA2000.DAT:249204 +			1432.00	2032.66	2.0e-105
/cgnl_8/gcgdata/geneseq/NA1999.DAT:V08864 +			794.50	1092.84	4.4e-53
/cgnl_8/gcgdata/geneseq/NA1999.DAT:V08874 +			794.50	1084.96	1.2e-52
/cgnl_8/gcgdata/geneseq/NA2000.DAT:A08690 +			631.50	884.50	1.8e-41
/cgnl_8/gcgdata/geneseq/NA2000.DAT:A08698 +			631.50	856.62	6.3e-40
/cgnl_8/gcgdata/geneseq/NA2000.DAT:249203 +			605.00	857.70	5.5e-40
/cgnl_8/gcgdata/geneseq/NA1990.DAT:004476 +			330.50	437.31	1.4e-16
/cgnl_8/gcgdata/geneseq/NA2000.DAT:A08697 +			324.00	423.81	5217
/cgnl_8/gcgdata/geneseq/NA2000.DAT:A08684 +			323.00	434.88	2.0e-16
/cgnl_8/gcgdata/geneseq/NA1994.DAT:V11546 +			315.00	408.55	6384
/cgnl_8/gcgdata/geneseq/NA1999.DAT:V99284 +			290.00	369.04	9.1e-13
/cgnl_8/gcgdata/geneseq/NA1999.DAT:220845 +			280.00	358.25	3.6e-12
/cgnl_8/gcgdata/geneseq/NA1995.DAT:095155 +			278.00	355.53	5.1e-12
/cgnl_8/gcgdata/geneseq/NA1999.DAT:220844 +			278.00	355.47	5.2e-12
/cgnl_8/gcgdata/geneseq/NA1999.DAT:220846 +			278.00	355.43	5.2e-12
/cgnl_8/gcgdata/geneseq/NA1991.DAT:Q12188 +			277.00	354.85	5.6e-12
/cgnl_8/gcgdata/geneseq/NA1999.DAT:220842 +			277.00	354.10	6.2e-12
/cgnl_8/gcgdata/geneseq/NA1999.DAT:220843 +			277.00	354.01	6.3e-12
/cgnl_8/gcgdata/geneseq/NA1993.DAT:038106 +			277.00	353.73	6.5e-12
/cgnl_8/gcgdata/geneseq/NA2000.DAT:247490 +			276.00	352.03	8.1e-12
/cgnl_8/gcgdata/geneseq/NA2000.DAT:247489 +			276.00	351.48	8.7e-12
/cgnl_8/gcgdata/geneseq/NA2000.DAT:247496 +			276.00	347.30	1.5e-11
/cgnl_8/gcgdata/geneseq/NA2000.DAT:247493 +			276.00	347.22	1.5e-11
/cgnl_8/gcgdata/geneseq/NA2000.DAT:247494 +			276.00	347.14	1.5e-11
/cgnl_8/gcgdata/geneseq/NA2000.DAT:247495 +			276.00	347.10	1.5e-11
/cgnl_8/gcgdata/geneseq/NA2000.DAT:289190 +			275.00	348.71	1.2e-11
/cgnl_8/gcgdata/geneseq/NA1990.DAT:005751 +			274.50	358.05	3.7e-12
/cgnl_8/gcgdata/geneseq/NA1994.DAT:Q58522 +			263.00	336.22	6.1e-11
/cgnl_8/gcgdata/geneseq/NA1998.DAT:V22837 +			199.00	240.52	1.3e-05
/cgnl_8/gcgdata/geneseq/NA1996.DAT:T27394 +			187.50	224.57	0.0001
/cgnl_8/gcgdata/geneseq/NA1998.DAT:V71604 +			187.50	224.56	0.0001
/cgnl_8/gcgdata/geneseq/NA1999.DAT:200195 +			186.00	222.35	0.0001
/cgnl_8/gcgdata/geneseq/NA1999.DAT:200195 +			186.00	222.35	0.0001
/cgnl_8/gcgdata/geneseq/NA1998.DAT:V61687 +			185.00	225.01	9.6e-05
/cgnl_8/gcgdata/geneseq/NA1998.DAT:V61690 +			185.00	220.93	0.0002
/cgnl_8/gcgdata/geneseq/NA1994.DAT:Q45197 +			180.50	214.53	0.0004

/cgnl_8/gcgdata/geneseq/NA1999.DAT:200266 + 178.50 211.49 0.0005 7277
/cgnl_8/gcgdata/geneseq/NA1999.DAT:200267 + 178.50 211.49 0.0005 7277
/cgnl_8/gcgdata/geneseq/NA1999.DAT:X23031 + 174.00 209.16 0.0007 5127
/cgnl_8/gcgdata/geneseq/NA1999.DAT:X23032 + 174.00 205.16 0.0012 7207
/cgnl_8/gcgdata/geneseq/NA1998.DAT:V31271 - 173.50 201.42 0.0020 9319

seq_name: /cgnl_8/gcgdata/geneseq/NA2000.DAT:249208

seq_documentation_block:

ID	249208	standard; DNA; 6714 BP.
XX		
AC	249208;	
XX		
DT	07-MAR-2000	(first entry)
XX		
DE	GLRaV-3 ORF 1a, encoding polyprotein.	
XX		
KW	Grapevine; leafroll; grapevine leafroll-associated virus 3; GLRaV-3;	
KW	viral disease; yield loss; sugar content; inhibition; infection;	
KW	replication; polyprotein; domain; proteinase; methyltransferase;	
KW	helicase; RNA-dependent; RNA polymerase; untranslated region;	
KW	transgenic plant; component; resistant; truncation; deletion; antisense;	
KW	expression; detection; antibody; ds.	
XX		
OS	Grapevine leafroll-associated virus 3.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	1..6714
FT		/tag= a
FT		/note= "ORF 1a"
FT		/product= "GLRaV-3 polyprotein"
FT	mat_peptide	253..612
FT		/tag= b
FT		/product= "GLRaV-3 polyprotein proteinase domain
FT		(Y58143)"
FT	mat_peptide	1378..2193
FT		/tag= c
FT		/product= "GLRaV-3 polyprotein methyltransferase domain
FT		(Y58144)"
FT	mat_peptide	5764..6636
FT		/tag= d
FT		/product= "GLRaV-3 polyprotein helicase domain (Y58145)"
XX		
PN	WO9955880-A1.	
XX		
PD	04-NOV-1999.	
XX		
PF	29-APR-1999;	99WO-US09307.
XX		
PR	29-APR-1998;	98US-0083404.
XX		
PA	(CORR) CORNELL RES FOUND INC.	
XX		
PI	Gonsalves D, Ling K;	
DR	WPI; 2000-062035/05.	
XX	P-PSDB; Y58148.	
XX		
PT	Newly isolated grapevine leafroll virus protein or polypeptide useful	
XX	for producing transgenic plants conferring viral disease resistance -	
XX	Claim 7; Fig 4; 84pp; English.	
PS		
XX		
CC	This sequence represents open reading frame (ORF) 1a which encodes	
CC	the polyprotein of the grapevine leafroll-associated virus 3	
CC	(GLRaV-3). Leafroll is a serious viral disease, occurring wherever	
CC	grapes are grown. Although not lethal, it causes yield losses and	
CC	reduction in sugar content. The virus encodes several proteins,	
CC	which may serve as targets for the inhibition of viral infection or	
CC	replication. These proteins include the 242-248 kD polyprotein (Y58148),	
CC	encoded by open reading frame (ORF) 1a, and which comprises a proteinase	
CC	domain (Y58143), a methyltransferase domain (Y58144) and a helicase	
CC	domain (Y58145); a RNA-dependent RNA polymerase (Y58146), encoded by ORF	

501 PheSerArgArgCysGlyAsnValCysPhePheAspIleGlySerPh 517
1501 TTTTCCAGGAGGTGGTAAATGTCCTTCTTCGATATTGGGGGAGCTT 1550
517 eThrTyrHisValLysAlaGlyHisValAsnCysHisValCysAsnProV 534
1551 CAGGTATCATGTCAAAAGCTGGCCATGTGAACGTCAATGATGATCAATCCAG 1600
534 aLeuAspValLysAspValLysArgArgGileAsnGluIleLeuPheLeu 550
1601 TCTAGACGTTAAAGATGTGAAGCGGAGAAATCAATGAGATCCTCTTCTT 1650
551 SerThrAlaGlyCysPheSerTyrValSerSerAspLeuLeuThrGluAl 567
1651 TCCACAGCTGGGGGAGATTCCGTACGTGTCACGTGACCTTCTAACTGAAGC 1700
567 aAlaSerLysSerValSerTyrCysSerArgGluSerGlnAsnCysAspS 584
1701 GGCTTCAAGTCTGTCTTACTGTAGTCGAGNATCGCAGNACTGCGATT 1750
584 eArgAlaAspAlaGlyPheMetValAspValTyrAspIleSerProGln 600
1751 CTAGAGCCGATCGGGTTTTATGGTGGATGTGTACGATATATCCCCGCGAG 1800
601 GlnValAlaGluAlaMetAspLysLysGlyAlaLeuValPheAspIleAl 617
1801 CAGGTAGCAGACGCTATGGATAAAGAGGGTGGCGTGGTTTTGCACATAGC 1850
617 aLeuMetPheProValGluLeuLeuTyrGlyAsnGlyGluValTyrLeuG 634
1851 TCTTATGTTCCCGTGGAGGTGTTGTACGGTACGGTGAAGTTTACTTGG 1900
634 luGluLeuAspThrLeuValLysArgGluGlyAspTyrLeuAlaTyrAsn 650
1901 AAGAACTCGATACGTTGGTGAAGAGGAAGGTGATTACCTGGCCATACAAT 1950
651 ValGlyGlnCysGlyGluMetTyrGluHisSerPheSerAsnValSerGl 667
1951 GTTGGTCACTGTGGTGAGATGTATGAACATTCCTTCTTAACGTAAGCCG 2000
667 yPhePheThrPheSerTyrValArgThrSerSerGlyAsnValPheLysL 684
2001 GTTTTTCACCTTTTCTATGTAGCCACTTCGTCCGGGAACGTGTTTAAGC 2050
684 euGluTyrGluGlyTyrArgCysGlyTyrHisHisLeuThrMetCysArg 700
2051 TAGAGTATCAGGGATACCGTTGTGGTTACCATCATCTCATGTGTAGG 2100
701 AlaGlnLysSerProGlyThrGluValThrTyrArgSerLeuValProSe 717
2101 GCTCAGAACTACCTTGGAACTCAGGTTACGTATAGGTCTGTGTCGCCGTC 2150
717 rPheValGlyLysSerLeuValPheIleProValValAlaGlySerSerV 734
2151 GTTCGTGGCAAAATCCCTGGGTTCATACCTGTGTAGCTGGTTCTAGTG 2200
734 aIserPheLysThrIleValLeuAspSerAspPheValAspArgIleTyr 750
2201 TGTCTTTTAAGACAATAGTCTCGATTCCGACTTTCTCCACAGGATCTAT 2250
751 SerTyrAlaLeuAsnThrIleGlyThrPheGluAsnArgThrPheGluTy 767
2251 TCTACCGCTCAACACTATAGGACATTCGAGAAATAGAACGTTTCAGTA 2300
767 rAlaValGlyAlaValArgSerGlnLysThrHisValIleThrGlySerA 784
2301 TGGCGTTGGGGGGTTCAGGTCCGAAAGACCCTATTCATTACAGGGAGTC 2350
784 rgValValHisSerLysValAspIleSerProAspAspMetTrpGlyLeu 800
2351 GCGTGTCCACAGCAGGTTGATATTCTCTTCATGATATGTGGGGTTTA 2400

801 valValAlaValMetAlaGlnAlaIleLysAspArgAlaLysSerIleAr 817
2401 GTTGTCCGCTGTTATGGCTCAGGCGATTAAAGGATAGGCGGAAGATATTCG 2450
817 gSerTyrAsnPheIleLysAlaSerGluGlySerLeuAlaGlyValPheL 834
2451 CTCCTATACTTTATAAAGCCAGTGAAGGAGTCTCGCCGGGGTCTTCA 2500
834 ysLeuPhePheGlnThrValLgIlyAspCysPheSerAsnAlaValSerVal 850
2501 AGCTCTTCTTTTCAGACCGTAGGCGATTGTTTTTCGAACGCGAGTCTCCGTC 2550
851 TyrAlaLysAlaMetValHisAspAsnPheAsnValLeuGluThrLeuMe 867
2551 TATGCTAAAGCAATGGTGCACGATAACTTCAACGTTTTTGAGACGCTTAT 2600
867 tSerMetProArgAlaPheIleArgLysValProGlySerValValValT 884
2601 GTCTATGCCAGAGCGTTCACTCCGTAAAGTACCTGGGTCTGTTGTTGTTA 2650
884 hrIleCysThrSerGlyAlaSerAspArgLeuGluLeuArgGlyAlaPhe 900
2651 CCATTTGCACCTTCTGGAGCTTCAGACAGGTTGGAGCTCAGGGGTGCCTTT 2700
901 AspIleSerLysGluThrPheGlyArgLysLeuLysAsnSerArgLeuAr 917
2701 GATATTTTGAAGGAGACCTTCGGTAGGAACCTGAAGAAATAGTCGCTGCG 2750
917 gValPheSerArgAlaIleValGluAspSerIleLysValMetLysAlam 934
2751 CGTCTTCTTAGGGCTATCGTGAAGATTCAATTAAGGTCATGAAGCAA 2800
934 etLysThrGluAspGlyLysProLeuProIleThrGluAspSerValTyr 950
2801 TGAAGACAGAAGATGGAACCCCTGCCAATTACTGAAGATTCTGTATAT 2850
951 AlaPheIleMetGlyAsnValSerAsnValHisCysThrArgAlaGlyLe 967
2851 GCCTTCATATATGGGAACGTTTCTAACGTCACGTACGAGGCGAGTCT 2900
967 uLeuGlyLysSerLysAlaThrValValSerSerValSerLysGlyLeuV 984
2901 TCTTGGCGTTCCAAAGCGACCGCTGGTTTCGAGTGTTCCTAAGGGTTGG 2950
984 aAlaArgGlyAlaAlaThrLysAlaPheSerGlyIleThrSerPhePhe 1000
2951 TAGCTCGTGGGGCTGCGACGAGGCGCTTTCTGSCATTACGTCGTTCTTT 3000
1001 SerThrGlySerLeuPheTyrAspArgGlyLeuThrGluAspGluArgLe 1017
3001 TCCACAGGTTCACTATTCTACGACCGCGGTTTAACTGAAGATGAAAGGCT 3050
1017 uAspAlaLeuValArgThrGluAsnAlaIleAsnSerProValGlyIleL 1034
3051 TGATGCTCTGGTGGCCACAGAGAAATGCTATAAACTCACCGGTGGGCATAC 3100
1034 euGluThrSerArgValAlaValSerLysValValAlaGlyThrLysGlu 1050
3101 TCGAGACGTCGCGCGTAGCTGTGAGCAAGCTGCTAGCTGGAACCAAGAA 3150
1051 PheTrpSerGluValSerLeuAsnAspPheThrThrPheValLeuArgAs 1067
3151 TTTTGGAGTGAAGTTTCTTAAATGACTTCCACACTTTTCGTATTTCGGAA 3200
1067 nLysValLeuIleGlyIlePheValAlaSerLeuGlyAlaAlaProIleA 1084
3201 TAAGGTGCTTATCGGATATTCGTGGCGCTTTTGGGTGCGGCCCAATTCG 3250
1084 latrPlystTyrArgArgGlyIleAlaAlaAsnAlaArgArgTyrAlaGly 1100
3251 CATGGAAGTATAGCGCGGAATTTCCGGCTAACCGCTAGAAGGTACGCGGC 3300
1101 SerSerTyrGluThrLeuSerSerLeuSerGlnAlaAlaGlyGlyLe 1117

||||| 3301 AGTAGTTACGAAACTCTAAGCTCGTTAAGTTTCACAAGCCGCGGTGGTTT 3350
||||| 1117 uArgGlyLeuThrSerSerThrValSerGlyGlySerLeuValValArgA 1134
||||| 3351 ACCCGGTTTAACTCTACGACACATTCGGTGGATCTTTAGTCGTCCGAA 3400
||||| 1134 rGlyPheSerSerAlaValThrValThrArgAlaThrValAlaLysArg 1150
||||| 3401 GAGGGTTTTGCTCGCGGTGACCGTCACTAGGGCGACGTAGCTAAACGT 3450
||||| 1151 GlnValProLeuAlaLeuLeuSerPheSerThrSerTyrAlaIleSerGI 1167
||||| 3451 CAAGTCCCTTTAGCGTTGCTATCGTTTTCTACCTCATACGCCATTTCCGG 3500
||||| 1167 yCysSerMetLeuGlyIleTTPAlaHisAlaLeuProArgHisLeuMetP 1184
||||| 3501 CTGCACATGATTAGGCAATTTGGGCACATGCTCTCCACGGCACATTAATGT 3550
||||| 1184 hePhePheGlyLeuGlyThrLeuLeuGlyAlaArgAlaSerAlaAsnThr 1200
||||| 3551 TTTTCTTTGGTTAGGGACATTTGCTTTGGGCGAGGGCTAGCGGAATACT 3600
||||| 1201 TrpLysPheGlyGlyPheSerAsnAsnTrpCysAlaValProGluValVa 1217
||||| 3601 TGGAAAGTTTGGAGGCTTCTCCAATAATTTGGTCCGCTGTTCCCGAGGTTGT 3650
||||| 1217 LtrArgGlyLysSerValSerLeuLeuLeuProIleThrLeuGlyV 1234
||||| 3651 TTGGCCAGGGAAGAGTGTCACTCATTTGTTACTGCCATTATACGTAGGGG 3700
||||| 1234 alSerLeuIleIleArgGlyLeuLeuAsnAspThrIleProGlnLeuAla 1250
||||| 3701 TATCTTTGATCATAGGGCTTGCTTAAACGACACCATACCTCAACTTGT 3750
||||| 1251 TyrValProProValGluGlyArgAsnValTyrAspGluThrLeuArgTy 1267
||||| 3751 TACGTCCACCGGTAGAGGGGAGGAATGTACGATGAGACGCTTAGGTA 3800
||||| 1267 rTyrArgAspPheAspTyrAspGluGlyAlaGlyProSerGlyThrGlnH 1284
||||| 3801 TTACCGGGACTTTGACTATGACGAAGGTGCTGGTCCATCTCGGACTCAGC 3850
||||| 1284 IsGluAlaValProGlyAspAsnAspGlySerThrSerSerValSer 1300
||||| 3851 ATGAAGCGGTTCCCGGTGACGATAACGATGGATCCACTTCTAGTGTCTCA 3900
||||| 1301 SerTyrAspValValThrAsnValArgAspValGlyIleSerThrAsnGI 1317
||||| 3901 AGCTATGATGTTGTACAAATGTGCGCGACGTGGGGATTAGCACCAACGG 3950
||||| 1317 yGluValThrGlyGluGluGluThrHisSerProArgSerValGlnTyrT 1334
||||| 3951 GGAAGTTACTGTTGGAAGAAGAGACCCATTACCTCCGAACCGTGCATACA 4000
||||| 1334 hrTyrValIgluGluGluValAlaProSerAlaAlaValAlaGluArgGln 1350
||||| 4001 CTTATGTCGAGGAAGAGGTTGCCCGTCTGCAGCTGTGGCGGGAAGACAA 4050
||||| 1351 GlyAspProSerGlySerGlyThrAlaAspAlaMetAlaPheValGluSe 1367
||||| 4051 GGTGATCCGTCGGGTTCTGTTACCGCTGACGCTATGGCTTTTGTGAAAG 4100
||||| 1367 rValLysLysGlyValAspAspValPheHisGlnGlnSerSerGlyGluT 1384
||||| 4101 TGTGAAAAAAGGTGTCGACGATGCTTTCCACCAACAGTCTAGTGGGAAA 4150
||||| 1384 hrAlaArgGluValGluValAlaAspGlyLysGlyLeuLeuProGluSerVal 1400
||||| 4151 CGGCTCGTGAGGTTGAGGTGACGCGCAAGGGTTGCTCCCGAAGACGTC 4200
||||| 1401 ValGlyGluAlaProThrGlnGluArgGlyArgAlaAlaAspGlyAsnTh 1417
|||||

4201 GTCCGTGAGGGCCCGACACACAGAAAGGGGAGAGCTGCAGATGGTAACAC 4250
||||| 1417 rAlaGlnThrAlaValAsnGluGlyAspArgGluProValGlnSerSerL 1434
||||| 4251 AGCACAACCCGGGTCAAGGAAGGCGACAGGAGCCAGTACAGTCCAGTC 4300
||||| 1434 euValSerSerProGlnAlaAspIleProLysValThrGlnSerGluVal 1450
||||| 4301 TTGTGAGTTCCGCACAGCTGATATTCCAAGGTCACCCAGTCCGAGGTA 4350
||||| 1451 HisAlaGlnLysGluValLysGlnGluValProLeuAlaThrValSerGI 1467
||||| 4351 CATGCTCAGAAAGAGTGAACAAGAGTACCATTGGCGACTGTTTCGGG 4400
||||| 1467 yAlaThrProIleValAspGluLysProAlaProSerValThrThrArgG 1484
||||| 4401 CGCCACGCCAATCGTCGATGAGAAACCCGCCCAAGTGTAGCACTCGTG 4450
||||| 1484 lyValLysIleIleAspLysGlyLysAlaValAlaHisValAlaGluLys 1500
||||| 4451 GTGTGAAGATAATTGACAAGGCAAGGCGTGCCTCATGTGGCTCAGAAA 4500
||||| 1501 LysGlnValGlnValGluGlnProLysGlnArgSerLeuThrIleAsnGI 1517
||||| 4501 AAACAGGTACAAAGTCGACGACGCCCAACAGAGGAGTTTGACGATCAATGA 4550
||||| 1517 uGlyLysAlaGlyLysGlnLeuCysMetPheArgThrCysSerCysGlyV 1534
||||| 4551 AGGCNAGCCCGGTAAACAGCTTGTGATGTTTAGAACGTGTTCTCGCGTG 4600
||||| 1534 alGlnLeuAspValTyrAsnGluAlaThrIleAlaThrArgPheSerAsn 1550
||||| 4601 TGCAGCTGGATGTGTACAACGAAGGAGCTATCGCCACCAGGTTCTCAAAC 4650
||||| 1551 AlaPheThrPheValAspAsnLeuLysGlyArgSerAlaValPhePheSe 1567
||||| 4651 GCATTTTACCTTTGTGATACTTGAAGGGAGGAGTGGCGTCTTTTTC 4700
||||| 1567 rLysLeuGlyGluGlyTyrThrTyrAsnGlyGlySerHisValSerSerG 1584
||||| 4701 AAAGCTGGGTGAGGGGTATACCTATATGTTGGTAGCCATGTTTCATCAG 4750
||||| 1584 lyTrpProArgAlaLeuGluAspIleLeuThrAlaIleLysTyrProSer 1600
||||| 4751 GGTGGCTCGTCCCTAGAGGATATCTTAACGGCAATTAAGTACCACAAGC 4800
||||| 1601 ValPheAspHisCysLeuValGlnLysTyrLysMetGlyGlyValPr 1617
||||| 4801 GTCTTCGACCACCTGTTTAGTGCAGAAGTACAAGATGGGTGGAGCGTACC 4850
||||| 1617 oPheHisAlaAspAspGluGluCysTyrProSerAspAsnProIleLeuT 1634
||||| 4851 ATTCACCGCTGATGACGAGGAGTGTATCCATCAGTAAACCTATCTTGA 4900
||||| 1634 hrValAsnLeuValGlyLysAlaAsnPheSerThrLysCysArgLysGly 1650
||||| 4901 CGGTCAATCTCGTGGGGAAGGCAAACTTCTCGACTAAGTGCAGGAAGGT 4950
||||| 1651 GlyLysValMetValIleAsnValAlaSerGlyAspTyrPheLeuMetPr 1667
||||| 4951 GGTAAAGTCACTGGTCATAAAGCTAGCTTTCGGGTGACTATTTTCTTATGCC 5000
||||| 1667 oCysGlyPheGlnArgThrHisLeuHisSerValAsnSerIleAspGluG 1684
||||| 5001 TTGCGGTTTTCAAGGACGCACTTGCATTCAGTAAACTCCATCCAGCAGAG 5050
||||| 1684 lyArgIleSerLeuThrPheArgAlaThrArgValPheGlyValGly 1700
||||| 5051 GCGCATCAGTTTGACGTTTACGGGCAACTCGCGCGCTCTTTGGTGTAGGC 5100
||||| 1701 ArgMetLeuGlnLeuAlaGlyValSerAspGluLysSerProGlyVa 1717
||||| 5101 AGGATGTTGCAGTTAGCCGCGCGGTGTCCGATGAGAAAGTCAACAGGTGT 5150

```
1717 lProAsnGlnGlnProGlnSerGlnGlyAlaThrArgThrIleThrProL 1734
|||||
5151 TCCAAACCCAGCAACCCAGAGCCAAAGGTGCTACCCAGAACAAATCACACCAA 5200
|||||
1734 ySerGlyGlyLysAlaLeuSerGluGlySerGlyArgGluValValGly 1750
|||||
5201 AATCGGGGGCAAGGCTCTATCTGAGGGAAGTGGTAGGGAAGTCAAGGGG 5250
|||||
1751 ArgSerThrTyrSerIleTrpCysGluGlnAspTyrValArgLysCysG 1767
|||||
5251 AGCTCGACATACCTCGATATGCTGCGCAACAAGATTACCTTAGGAAGTGTGA 5300
|||||
1767 uTrpLeuArgAlaAspAsnProValMetAlaLeuGluProAspTyrThr 1784
|||||
5301 GTGGCTCAGGGCTGATAATCCAGTGTGCTTGAACCTGACTACACCC 5350
|||||
1784 roMetThrPheGluValValLysThrGlyThrSerGluAspAlaValVal 1800
|||||
5351 CAATGACATTGGAAGTGGTTAAACCCGGGACCTCTGAAGATGCCGTCGTG 5400
|||||
1801 GluTyrLeuLysTyrLeuAlaIleGlyIleGluArgThrTyrArgAlaLe 1817
|||||
5401 GAGTACTTCAAGTATCTGGCTATAGGCATTGACAGGACATACAGGCGTT 5450
|||||
1817 uLeuMetAlaArgAsnIleAlaValThrThrAlaGluGlyValLeuLysV 1834
|||||
5451 GCTTATGGCTAGAAATATTGCCGCTACTACCCGCCGAAGGTGTTCTGAAG 5500
|||||
1834 alProAsnGlnValTyrGluSerLeuProGlyPheHisValTyrLysSer 1850
|||||
5501 TACCTAATCAAGTTTATGAATCACTACCCGGGCTTTCACGTTTACAAGTCG 5550
|||||
1851 GlyThrAspLeuIlePheHisSerThrGlnAspGlyLeuArgValArgAs 1867
|||||
5551 GGCACAGATCTCATTTTTCATTCAACACAAAGACGGCTTGCGTGTGAGAGA 5600
|||||
1867 pLeuProTyrValLeuIleAlaGluLysGlyIlePheThrLysGlyLysA 1884
|||||
5601 CCTACCGTAGTACTCATAGCTGAAAGAGTATCTTTACCAAGGGCAAG 5650
|||||
1884 spValAspAlaValAlaLeuGlyAspAsnLeuPheValCysAspAsp 1900
|||||
5651 ATGTCACCGCGGTGGTAGCTTTGGGCGACAAATCTGTTCGTATCGGACGAT 5700
|||||
1901 ileLeuValPheHisAspAlaIleAsnLeuIleGlyAlaLeuLysValAl 1917
|||||
5701 ATACTGGTTTTCCACGATGCCATTATTTGCATAGGTGCACCTGAAAGTCGC 5750
|||||
1917 aArgCysGlyMetValGlySerPheLysSerPheGluTyrLysCysT 1934
|||||
5751 TCGATCGGCATGGTGGGCAATTCGTTTAAGTCTCTCGAATATAAGTGCT 5800
|||||
1934 yrAsnAlaProProGlyGlyGlyLysThrThrThrLeuValAspGluPhe 1950
|||||
5801 ATAATGCTCCCCCGGTGGCGGTAAACGACGACGCTTAGTGGACGAATC 5850
|||||
1951 ValLysSerProAsnSerThrAlaThrIleThrAlaAsnValGlySerSe 1967
|||||
5851 GTTAAAGTCACCAATAGCAGACCCACCATTTACGGCTAATGTGGGAAGTTC 5900
|||||
1967 rGluAspIleAsnMetAlaValLysLysArgAspProAsnLeuGluGlyL 1984
|||||
5901 TGAGGACATAAATATGGCGGTGAAGAGAGATCCGGAATTTTGGAAAGTTC 5950
|||||
1984 euAsnSerAlaThrThrValAsnSerArgValValAsnPheIleValArg 2000
|||||
5951 TCAAAGTGTCTACCAAGTTAACTCCAGGCTGTGTAACCTTTATCGTCAGG 6000
|||||
2001 GlyMetTyrLysArgValLeuValAspGluValHisMetMetHisGlnG 2017
|||||
6001 CGAATCTATAAAGGGTTTGGTGGATAGGTGCACATGATGATCAAGG 6050
|||||
```

```
2017 yLeuLeuGlnLeuGlyValPheAlaThrGlyAlaSerGluGlyLeuPheP 2034
|||||
6051 CTTACTACAACTAGGCGCTTCGCACCCGCGCGCTCGAAGGCCCTCTTTT 6100
|||||
2034 heGlyAspIleAsnGlnIleProPheIleAsnArgGluLysValPheArg 2050
|||||
6101 TTGGAGACATAAATCAGATACCATTTCAAAACAGGGAGAGGTGTTTAGG 6150
|||||
2051 MetAspCysAlaValPheValProLysLysGluSerValValTyrThrSe 2067
|||||
6151 ATGGATTGCTGCTGTTTGTTCCAAAGAAGAAAGCGTTGTATACACTTC 6200
|||||
2067 rLysSerTyrArgCysProLeuAspValCysTyrLeuLeuSerSerMetT 2084
|||||
6201 TAAATCGTACAGGTGCTCCGTAGATGTTGCTACTTGTGCTCTCAATGA 6250
|||||
2084 hrValArgGlyThrGluLysCysTyrProGluLysValValSerGlyLys 2100
|||||
6251 CCGTAAGGGGAACGGAAAGTGTACCTCTGAAAGGTCGTTAGCGGTAAG 6300
|||||
2101 AspLysProValValArgSerLeuSerLysArgProIleGlyThrThrAs 2117
|||||
6301 GACAAACCCAGTAGTAAGATCGCTGTCCAAAAGGCCAATTGGAACCACTGA 6350
|||||
2117 pAspValAlaGluIleAsnAlaAspValTyrLeuCysMetThrGlnLeuG 2134
|||||
6351 TGACGTAGCTGAAATAAACGCTGACGCTGACTTGTGTCATGACCCAGTTGG 6400
|||||
2134 LuLysSerAspMetLysArgSerLeuLysGlyLysGlyLysGluThrPro 2150
|||||
6401 AGAAGTCGGATATGACAGGCTGTTGAAGGGAAGAAAGAAACACCA 6450
|||||
2151 ValMetThrValHisGluAlaGlnGlyLysThrPheSerAspValValLe 2167
|||||
6451 GTGATGACAGCTGCATGAAGCACACAGGGAACAACTTCAGTGATGTGGTATT 6500
|||||
2167 uPheArgThrLysLysAlaAspAspSerLeuPheThrLysGlnProHisI 2184
|||||
6501 GTTTAGGACGAAGAAGCCGATGACTCCCTATTTCACATAAACACCCGATA 6550
|||||
2184 leLeuValGlyLeuSerArgHisThrArgSerLeuValTyrAlaAlaLeu 2200
|||||
6551 TACTTGTGTTGTTGTCGAGACACACGCTACCTGCTTTATGCCGCTCTG 6600
|||||
2201 SerSerLysLeuAspAspLysValGlyThrTyrIleSerAspAlaSerPr 2217
|||||
6601 AGCTCAAAGTTGGACGATAAGTTCGCGCACATATATTAGCGACGCGTCACC 6650
|||||
2217 oGlnSerValSerAspAlaLeuLeuHisThrPheAlaProAlaGlyCysP 2234
|||||
6651 TCAATCAGTATCGACGCTTTGCTTTCACAGCTTCGCCCGCGCTGGTGCT 6700
|||||
2234 heArgGlyIle 2237
|||||
6701 TTCAGGTATA 6711
|||||
seq_name: /cgnl_8/gcgdata/geneseq/geneseq/NA2000.DAT:249200
seq_documentation_block:
ID 249200 standard; DNA; 17919 BP.
XX
AC 249200;
XX
XX 07-MAR-2000 (first entry)
XX
DE Grapevine leafroll-associated virus 3 (GLRaV-3) genome.
KW Grapevine; leafroll; grapevine leafroll-associated virus 3; GLRaV-3;
KW viral disease; yield loss; sugar content; inhibition; infection;
KW replication; polyprotein; domain; proteinase; methyltransferase;
KW helicase; RNA-dependent; RNA polymerase; untranslated region;
KW transgenic plant; component; resistant; truncation; deletion; antisense;
KW expression; detection; antibody; ds.
```

```

XX OS Grapevine leafroll-associated virus 3.
XX FH Key Location/Qualifiers
XX FT 5'UTR 1..158
XX FT CDS 159..6872
XX FT /*tag- a
XX FT /*tag- b
XX FT /note- "ORF 1a"
XX FT /product- "GLRaV-3 polyprotein (Y58148)"
XX FT 411..770
XX FT /*tag- c
XX FT /product- "GLRaV-3 polyprotein proteinase domain
XX FT (Y58143)"
XX FT mat_peptide
XX FT 1536..2351
XX FT /*tag- d
XX FT /product- "GLRaV-3 polyprotein methyltransferase domain
XX FT (Y58144)"
XX FT mat_peptide
XX FT 5922..6794
XX FT /*tag- e
XX FT /product- "GLRaV-3 polyprotein helicase domain (Y58145)"
XX FT 6877..8478
XX FT /*tag- f
XX FT /note- "ORF 1b"
XX FT /product- "GLRaV-3 RNA-dependent RNA polymerase (Y58146)"
XX FT 8708..8863
XX FT /*tag- g
XX FT /note- "ORF 2"
XX FT 9930..10067
XX FT /*tag- h
XX FT /note- "ORF 3"
XX FT 10086..11735
XX FT /*tag- i
XX FT /note- "ORF 4"
XX FT 11728..13179
XX FT /*tag- j
XX FT /note- "ORF 5"
XX FT 13269..14210
XX FT /*tag- k
XX FT /note- "ORF 6"
XX FT 14273..15706
XX FT /*tag- l
XX FT /note- "ORF 7"
XX FT 15717..16274
XX FT /*tag- m
XX FT /note- "ORF 8"
XX FT 16271..16804
XX FT /*tag- n
XX FT /note- "ORF 9"
XX FT 16811..17350
XX FT /*tag- o
XX FT /note- "ORF 10"
XX FT 17353..17463
XX FT /*tag- p
XX FT /note- "ORF 11"
XX FT /product- "GLRaV-3 ORF 11-encoded protein (Y58147)"
XX FT 17460..17642
XX FT /*tag- q
XX FT /note- "ORF 12"
XX FT 17643..17919
XX FT /*tag- r
XX PN WO9955880-A1.
XX XX
XX PD 04-NOV-1999.
XX PF 29-APR-1999; 99WO-US09307.
XX XX
XX PR 29-APR-1998; 98US-0083404.
XX XX
XX PA (CORR ) CORNELL RES FOUND INC.
XX XX
XX PI Consalves D, Ling K;

```

```

XX WPI; 2000-062035/05.
XX DR P-PSDB; Y58143, Y58144, Y58145, Y58146, Y58147, Y58148.
XX XX
XX PT Newly isolated grapevine leafroll virus protein or polypeptide useful
XX PT for producing transgenic plants conferring viral disease resistance -
XX PS Example 1; Fig 2; 84pp; English.
XX XX
XX CC This sequence represents the genome of grapevine leafroll-associated
XX CC virus 3 (GLRaV-3). Leafroll is a serious viral disease, occurring
XX CC wherever grapes are grown. Although not lethal, it causes yield losses
XX CC and reduction in sugar content. The virus encodes several proteins,
XX CC which may serve as targets for the inhibition of viral infection or
XX CC replication. These proteins include a 242-248 kD polyprotein (Y58148),
XX CC encoded by open reading frame (ORF) 1a, and which comprises a proteinase
XX CC domain (Y58143), a methyltransferase domain (Y58144) and a helicase
XX CC domain (Y58145); a RNA-dependent RNA polymerase (Y58146), encoded by ORF
XX CC 1b; and a protein of unspecified function (Y58147), encoded by ORF 11.
XX CC Nucleotides encoding these proteins, or fragments thereof, and the 5',
XX CC and 3' untranslated regions (UTRs) of the genome (249201-249202) are
XX CC useful for the generation of transgenic plants and plant components.
XX CC Such transgenic plants may be resistant to viral disease, for example,
XX CC this property being conferred on the plants via the use of nucleotides
XX CC encoding truncated or internally deleted proteins, or via the use of
XX CC antisense nucleotides to inhibit viral gene expression. The nucleotides
XX CC may additionally be used in the detection of viral nucleic acids in a
XX CC tissue sample. The proteins can be used to generate antibodies which
XX CC may be used to detect GLRaV-3 in plant samples. The isolation of
XX CC GLRaV-3 facilitates the production of agents that reduce the risk of
XX CC infection or damage by the virus in vineyards.
XX SQ Sequence 17919 BP; 4729 A; 3527 C; 4740 G; 4923 T; 0 other;

alignment_scores:
    Quality: 11520.00      Length: 2237
    Ratio: 5.150          Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-09-301-906-15 x 249200 ..

Align seg 1/1 to: 249200 from: 1 to: 17919

1 MetAspTyrIleArgProLeuArgValPheSerPheProHisValAsnAs 17
159 ATGGACTACATTGCGCCCATTCGCGGTTTTCTCTTCTCCTCAGCTTAATAA 208

17 nThrLeuGluTyrValArgTyrAsnLysAlaAsnGlyAspValGlyAlap 34
209 CACCTTGGAGTAGCTAGGTACACAAGGCCAATGGTGATGTAGGAGCTT 258

34 heLeuThrThrMetLysPheIleGlyAsnValLysLeuSerAspPheThr 50
259 TCCTACCCACCACCATGAAGTTCATAGGAGACGTGAAGTTGTCGGACTTCACA 308

51 ProArgCysAlaAlaMetIleTyrIleGlyLysLeuThrLysGlyVally 67
309 CCCAGGTGCGCCACCTATGATTTACATTTGAAAAGCTCACCAAGGGGTGAA 358

67 sArgThrPheValProProValLysGlyPheAlaArgGlnTyrAlav 84
359 GCGTACGTTTGTCCCCACCACAGTTAAAGGGTTTGCACGGCAGTACGCTG 408

84 alValSerGlySerValSerAlaLeuArgGlyAspGlyLysLysValLeu 100
409 TTGTCAGCGGCTCAGTCAGCGGCTGAGAGGGGATGGTAAGAAGGCTCTTG 458

101 MetGluAlaArgThrSerThrSerAlaThrSerAspValSerAspPheAs 117
459 ATGGAGGCAAGGACCTCAACTTCCGCAACTTCCGACGTGCTGATTTCGA 508

```

117 pValValPheGluAlaValSerAsnAlaLeuLeuValValHisTyrHisA 134
|||||
509 CGTCGATTGGAAGCTGTTCTTAATGCATTACTTGCGTACACTACCAC 558
134 rgValValProTyrAlaProValValArgGluGlnProLysProAlaVal 150
|||||
559 GCGTAGTGGCGTATGCCCGCTCAAGCGGAGCAGCGCTAAACGGCTGTT 608
151 LysGlnAspGluGlnLysProLysArgGlnAlaSerHisTrpAlaValLy 167
|||||
609 AGCAAGATGAGCAGAAGCCCAACGCGAAGCGTCACATTGGGCTGTAA 658
167 sProThrAlaValGlyValHisValProLeuProLysLysGlnGluAla 184
|||||
659 GCCAACAGCTGTTGGCGTCCACGTACCACCTTCTTAAACACAGGAGCAC 708
184 euGluProAlaGlnSerValProGlnGlnSerLeuGluGluLysAlaAla 200
|||||
709 TGGAGCCAGCGCAATCAGTCCCAACAACAGTCGTTGGAGGAGAAGCGGCC 758
201 LeuThrPheGlyLeuPhePheSerLysGlyGlyAspGluSerAspAl 217
|||||
759 TTGACGTTTGGCCCTTTCTTCAGTAAGGTGGGGGTGCATGAGAGCGACGC 808
217 aValIleLeuArgLysGlyLysLeuPheAsnArgAlaLeuAsnValProI 234
|||||
809 TGTCACTTGGCGAAGAGGNAATGTTTACAGGGCCCTTAATGTTCCCTA 858
234 leAspValLysAsnThrPheValTrpAlaLysIleTrpAspGluAlaSer 250
|||||
859 TTGATGTAAAGAACACGTTCTGTTGGCTTAAATCTGGGATGAAGCCTCT 908
251 ArgArgArgGlyTyrPheTyrValLysAspArgAlaValLysPhePhePr 267
|||||
909 CPTAGGAGAGGGTATTTTACGTCAAGATAGAGCTGTTAAATCTTCCC 958
267 oIleValArgGlyArgAlaThrIleGluAspPheIleValAsnThrAlap 284
|||||
959 TATTGTGGGGGTAGGGCTACGATCAGAGACTTCATCGTAATACAGCCC 1008
284 roGlyCysAspValAlaLeuProArgIleGluLeuTrpSerMetArgGlu 300
|||||
1009 CAGGGTGTGATGTTGCTTGGCCGCAATTGAGTTGTGGAGTATGCCGAA 1058
301 ArgAlaPheValCysThrThrLysGlyTyrCysTrpPheAsnAsnGluAr 317
|||||
1059 AGGGCGTTTGTATGCCACCACCAAGGGTGCTGTGTTTAAACAATGAGAG 1108
317 gLeuArgGlyGluIleTyrArgArgArgCysPheSerSerPheSerI 334
|||||
1109 GCTGAGGGAGGAATTTACAGAGCTGCTGCTCTCATCTTCCCTTTTCGA 1158
334 leGlyPheLeuMetHisLeuGlyPheArgSerLeuLysValIleArgPhe 350
|||||
1159 TAGGTTTCTTGATGCACCTTGGCTTTAGATCGTTAAAGGTCAATTAGTTT 1208
351 AlaGlyThrAsnIleLeuHisMetProSerLeuAsnGluGluArgThrPh 367
|||||
1209 GCGGGCAGCAACACTACTACACATGCCATCCTCAATCAATGAAGCGTACCTT 1258
367 eGlyTTrpLysGlyValAspValTyrLeuProAsnValProLysThrAlaI 384
|||||
1259 TGGGTGGAGGGCGAGAGCTCTATTCTCCCAATGTCCCAAAACCGCTA 1308
384 leValAlaGlyAspArgTrpArgLeuGlyGlyGluIleLeuAlaSerVal 400
|||||
1309 TCGTCGCTGGCGATAGACACGGTTGGAGGGAGATCTTGGGCTCCGTC 1358
401 AlaAsnAlaLeuAsnGlnGluGluValTyrSerSerValValSerSerIl 417
|||||
1359 GCCAATGCCCTTAAATCAAGAGAGGGTCTATTTCATCGGTCGTTTCGAGTAT 1408
417 eThrAsnArgLeuValLeuArgAspGlnSerAlaLeuLeuSerHisLeuA 434
|||||

1409 CACCAATAGACTGGTATTAAAGGACCAATCGCATTCCTTTCCCATTTGG 1458
434 spThrLysLeuCysAspMetPheSerGlnArgAspAlaMetIleArgGlu 450
|||||
1459 ACACGAAATTTGTCGATATGTTTCTCAAAGGACGCAATGATTTCGGCAA 1508
451 LysProSerHisArgCysAspValPheLeuLysProArgGluArgGluLy 467
|||||
1509 AAACCTTCACATAGGTGCGATGTGTTTCTGAAGCCGCGAAAGGGAGAA 1558
467 sLeuArgGluLeuPheProGluLeuSerIleGlnPheSerAspSerValA 484
|||||
1559 GCTGAGGAACTCTTCCAGAGCTTTCGATACAGTCTCCGACTCGGTCA 1608
484 rgSerSerHisProPheAlaAsnAlaMetArgSerCysPheAsnGlyIle 500
|||||
1609 GGAGTAGTCACCCATTCGCTAATGCCCATCGGAGCTGTTTCAATGGAATC 1658
501 PheSerArgArgCysGlyAsnValCysPhePheAspIleGlyGlySerPh 517
|||||
1659 TTTTCCAGGAGGTGGTAAATGTGTCTTCTTCATATTGGGGGAGCTT 1708
517 eThrTyrHisValLysAlaGlyHisValAsnCysHisValCysAsnProv 534
|||||
1709 CACGTATCATGTCAAAGCTGGCATGTGAACTGTCTATGTATGCAATCCAG 1758
534 alLeuAspValLysAspValLysArgArgIleAsnGluIleLeuPheLeu 550
|||||
1759 TCCTAGACGTTAAAGATGTCAAGCGGAGAACTAATGAGATCCTCTTCTT 1808
551 SerThrAlaGlyGlyAspSerTyrValSerSerAspLeuLeuThrGluAl 567
|||||
1809 TCCACAGCTGGGGAGATTCTGACGTCTCCAGTGACCTTCTTAAGTGAAGC 1858
567 aAlaSerLysSerValSerTyrCysSerArgGluSerGlnAsnCysAsps 584
|||||
1859 GCGTTCAAAGTCTGTCTTACTGTAGTCGAGAATCCACAGACTGCCGATT 1908
584 erArgAlaAspAlaGlyPheMetValAspValTyrAspIleSerProGln 600
|||||
1909 CTAGAGCCGATCGCGGTTTATGTTGGTGGATGTTACGATATATCCCGCAG 1958
601 GlnValAlaGluAlaMetAspLysLysGlyAlaLeuValPheAspIleAl 617
|||||
1959 CAGGTAGCAGAGGCTATGGATAAGAGGGTCCGCTGGTTTTTCGACATAGC 2008
617 aLeuMetPheProValGluLeuLeuTyrGlyAsnGlyGluValTyrLeuG 634
|||||
2009 TCTTATGTTCCCGTGGAGTTGTTGACGGTAACGGTGAAGTTTACTTTGG 2058
634 luGluLeuAspThrLeuValLysArgGluGlyAspTyrLeuAlaTyrAsn 650
|||||
2059 AACAACCTCGATACGTTGGTCAAGAGGGAAGGTGATTACCTGGCCTACAAT 2108
651 ValGlyGlnCysGlyGluMetTyrGluHisSerPheSerAsnValSerGl 667
|||||
2109 GTTGGTCAGTGTGGTGAAGTGTATGAACATTCTCTTAACGTAAAGCGG 2158
667 yPhePheThrPheSerTyrValArgThrSerSerGlyAsnValPheLysL 684
|||||
2159 GTTTTTCACCTTTTCTTATGTACGCACTTCGTCGGGAACGTGTTTAAGC 2208
684 euGluTyrGluGlyTyrArgCysGlyTyrHisHisLeuThrMetCysArg 700
|||||
2209 TAGAGTATGAGGATACCGTGTGGTTACCATCATCTCACTATGTGTAGG 2258
701 AlaGlnLysSerProGlyThrGluValThrTyrArgSerLeuValProSe 717
|||||
2259 GCTCAGAAGTCAGCTGGAACTGAGGTTACGTATAGGTGCTTGGTCCGCTC 2308
717 rPheValGlyLysSerLeuValPheIleProValValAlaGlySerSerV 734
|||||

2309 GTTCGTGGCAAAATCGCTGGTGTTCATACCTGCTTGTAGCTGCTTCTAGTG 2358
734 aIserPheLysThrIleValLeuAspSerAspPheValAspArgIleYr 750
2359 TGTCTCTTAAGACATATAGTCTCGATTCGGACTTTGTCTGACAGGATCTAT 2408
751 SerTyrAlaLeuAsnThrIleGlyThrPheGluAsnArgThrPheGluTy 767
2409 TCCTACGGCGCTCAACACTATAGGACATTCGAGAAATAGAACGCTTTAGTA 2458
767 rAlaValGlyAlaValArgSerGlnLysThrHisValIleThrGlySerA 784
2459 TGCCCTTGGGGCGGTACAGTCGCAAAAGACCCATGCTATTACAGGGAGTC 2508
784 rGValValHisSerLysValAspIleSerProAspAspMetTyrGlyLeu 800
2509 GGGTTGTCACACAGCAAGGTTGNATATTTCTCTGATGATATGTGGGGTTA 2558
801 ValValAlaValMetAlaGlnAlaIleLysAspArgAlaLysSerIleAr 817
2559 GTTGTGCGTGTATGGCTCAGCGGATTAAGGATAGGGCGAAGAGTATTCCG 2608
817 gSerTyrAsnPheIleLysAlaSerGluGlySerLeuAlaGlyValPheL 834
2609 CTCCTATAACTTTATAAAGCCAGTGAGGGAGTCTCGCCGGGGTCTTCA 2658
834 ySLeuPhePheGlnThrValGlyAspCysPheSerAsnAlaValSerVal 850
2659 AGCTCTCTTTCAGACCGTAGCGGATTTGTTTTCGAACGCAGTCTCCGTC 2708
851 TyrAlaLysAlaMetValHisAspAsnProAsnValLeuGluThrLeuMe 867
2709 TATGCTAAGGCAATGGTGCACGATAAATCAACGTTTGGAGACGCTTAT 2758
867 tSerMetProArgAlaPheIleArgLysValProGlySerValValValT 884
2759 GTCTATGCCAGAGCGTTTATCCCGTAAAGTACCTCGGGTCTGTGTGTGTA 2808
884 hrIleCysThrSerGlyAlaSerAspArgLeuGluLeuArgGlyAlaPhe 900
2809 CCATTTGCACITCTCGAGCTTCAGACAGGTTGGAGCTCAGGGGTGCCIT 2858
901 AspIleSerLysGluThrPheGlyArgLysLeuLysAsnSerArgLeuAr 917
2859 GATATTTGGAAGAGACCTTCGGTAGGAACACTGAAGAATAGTCGCTTGGC 2908
917 qValPheSerArgAlaIleValGluAspSerIleLysValMetLysAlaM 934
2909 CGCTCTCTAGGGCTATCGTGAAGATTCAAATTAAGGTCATGAAGGCAA 2958
934 eLysThrGluAspGlyLysProLeuProIleThrGluAspSerValTyr 950
2959 TGAACACAGAGATGSAANACCCCTGCCAATTAATGAAGATTCGTATAT 3008
951 AlaPheIleMetGlyAsnValSerAsnValHisCysThrArgAlaGlyLe 967
3009 GCGTTCATAATGGGAACGTTTCTAATCCACTGTACGAGGCGAGGCT 3058
967 uLeuGlyGlySerLysAlaThrValValSerSerValSerLysGlyLeuV 984
3059 TCCTGGCGGTTGGAAGGACCGTGGTTTCGAGTGTTCCTAAGGGTTTGG 3108
984 aAlaArgGlyAlaAlaThrLysAlaPheSerGlyIleThrSerPhePhe 1000
3109 TAGCTCGTGGGCTCGGACGAGGCGCTTTCTGGCATTTACGTCGTTCTTT 3158
1001 SerThrGlySerLeuPheTyrAspArgGlyLeuThrGluAspGluArgLe 1017
3159 TCCACAGGTTCACTATTCTACGACCGCGGTTTAACTGAAGATGAAGGCT 3208
1017 uAspAlaLeuValArgThrGluAsnAlaIleAsnSerProValGlyIleL 1034
3209 TGATGCTCTGGTGGCGACAGAGAAATGCTATAAACTCACCGGTGGGCATAC 3258

1034 euGluThrSerArgValAlaValSerLysValValAlaGlyThrLysGlu 1050
3259 TGGACAGCTCGCGCGTGTAGCTGTGACCAAGTGTAGCTGGAACGAAGAA 3308
1051 PheTyrSerGluValSerLeuAsnAspPheThrThrPheValLeuArgAs 1067
3309 TTTTGGAGTGAAGTTTCCCTAAATGACTTCACCACTTTTCGTATTGCGAA 3358
1067 nLysValLeuIleGlyIlePheValAlaSerLeuGlyAlaAlaProIleA 1084
3359 TAAAGTGTCTTATCGGGATATTCGTGGCGCTTTTGGGTGGGCCCAATTC 3408
1084 IaTrpLysTyrArgArgGlyIleAlaAlaAsnAlaArgArgTyrAlaGly 1100
3409 CATGGAAGTATAGGCGCGGAATTCGGCTACGCTAGAGGTACGCGGC 3458
1101 SerSerTyrGluThrLeuSerSerLeuSerSerGlnAlaAlaGlyGlyLe 1117
3459 AGTAGTTACGAAACTCTAAGCTCGTTAAGTTCAAGCGCGCGGTGGTTT 3508
1117 uArgGlyLeuThrSerSerThrValSerGlyGlySerLeuValValArgA 1134
3509 ACGCGGTTTAACTCTAGCACAGTATCCGGTGGATCTTTTAGTCGTCGAA 3558
1134 rGlyPheSerSerAlaValThrValThrArgAlaThrValAlaLysArg 1150
3559 GAGGGTTTTCGTGGCGGTGACCGTCACTAGGGCGACCGTAGCTAAAGCT 3608
1151 GlnValProLeuAlaLeuLeuSerPheSerThrSerTyrAlaIleSerGl 1167
3609 CAAGTCCCTTAGCGTTGCTATCGTTTCTACCTCATACGCCATTTCCCG 3658
1167 yCysSerMetLeuGlyIleTrpAlaHisAlaLeuProArgHisLeuMetP 1184
3659 CTGCAAGTATGTTAGGCATTTGGGCACATGCTCTTCCAGGCACCTTAAGT 3708
1184 hePhePheGlyLeuGlyThrLeuLeuGlyAlaArgAlaSerAlaAsnThr 1200
3709 TTTTCTTTGGTTTAGGCATTTGCTTGGGCGAGGGCTAGCGCGAATACT 3758
1201 TrpLysPheGlyGlyPheSerAsnAsnTrpCysAlaValProGluValVa 1217
3759 TGGAAAGTTTGGAGGCTTCTCCAATAATTTGGTGGCTGTTCGCGAGGTTGT 3808
1217 tTrpArgGlyLysSerValSerLeuLeuLeuProIleThrLeuGlyV 1234
3809 TTGGCGAGGGAAGAGTGTACGCTCATTTGTTACTGCTATTACGCTAGGGG 3858
1234 aISerLeuIleIleArgGlyLeuLeuAsnAspThrIleProGlnLeuAla 1250
3859 TATCTTTGATCATAGGGCTTGTCTTAAGCACACCATACCTCAACTGCT 3908
1251 TyrValProProValGluGlyArgAsnValTyrAspGluThrLeuArgTy 1267
3909 TACGTCCCAACCGGTAGAGGGGAGGAATGTACGATGACACGCTTAGGTA 3958
1267 rTyrArgAspPheAspTyrAspGluGlyAlaGlyProSerGlyThrGlnH 1284
3959 TTACCGGGACTTTTGACTATGACGAAGGTGCTGCTCCATCTGGGACTCAGC 4008
1284 tSGLuAlaValProGlyAspAsnAspGlySerThrSerSerValSer 1300
4009 ATGAAGCGGTTCCCGGTGACGATACGATGATGATCCACTTCTAGTGTCTCA 4058
1301 SerTyrAspValValThrAsnValArgAspValGlyIleSerThrAsnGl 1317
4059 ACCTATGATGTTGTCAAAATGTGCGCGACGTGGGATTTAGCACCAACGG 4108
1317 yGluValThrGlyGluGluThrHisSerProArgSerValGlnTyrT 1334
4109 GGAAGTTACTGGTGAAGAAGAGACCACTTCACTCGAAGCGTGCAATACA 4158

1334 hrTyrValGluGluGluValAlaIleProSerAlaAlaValAlaGluArgGln 1350
|||||
4159 CTTATGTCAGAGAGAGGTGGCCCCGTCGACGCTGGCGGAAGACAA 4208
1351 GlyAspProSerGlySerGlyThrAlaAspAlaMetAlaPheValGluSe 1367
|||||
4209 GGTGATCCGTCGGGTCTCGTACCGCTGACGCTATGGCTTTTGTCAAAG 4258
1367 rValLysLysGlyValAspValPheHisGlnGlnSerSerGlyGluT 1384
|||||
4259 TGTGAAAAAGGTGTGACGATGCTTTTACCACACAGTCTAGTGGGAAA 4308
1384 hrAlaArgGluValGluValAspGlyLysGlyLeuLeuProGluSerVal 1400
|||||
4309 CGGCTCGTGGGTTGAGGTGGACGGCAAGGGTTGCTCCAGAAAGCGTC 4358
1401 ValGlyGluAlaProThrGlnGluArgGlyArgAlaAlaAspGlyAsnTh 1417
|||||
4359 GTCGGTGAAGCGCCGACACAGAAAGGGGACAGCTGCAGATGGTAACAC 4408
1417 rAlaGlnThrAlaValAsnGluGlyAspArgGluProValGlnSerSerL 1434
|||||
4409 AGCACAAACCGGGTGCACGAAGGGGACAGAGGAGCCAGTACAGTCCAGTC 4458
1434 euValSerSerProGlnAlaAspIleProLysValThrGlnSerGluVal 1450
|||||
4459 TTGTGAGTTCCGCACAGGCTGATATTCCAAAGGTCACCCAGTCCGAGGTA 4508
1451 HisAlaGlnLysGluValLysGlnGluValProLeuAlaThrValSerG 1467
|||||
4509 CATGCTCACAAAGAGTGAACAAAGAGTACCATTGGCGACTGTTTCGGG 4558
1467 yAlaThrProIleValAspGluLysProIleValProSerValThrArgG 1484
|||||
4559 CGCCACGCCAATCGTCGATGAGAAACCCGCCCAAGTGTACGACTCGTG 4608
1484 lyValLysIleIleAspLysGlyLysAlaValAlaHisValAlaGluLys 1500
|||||
4609 GTGTGAAGATAATTGACAAGGGCAAGCGTCGCTCATGTGGCTGAGAAA 4658
1501 LysGlnValGlnValGluGlnProLysGlnArgSerLeuThrIleAsnG 1517
|||||
4659 AAACAGGTACAAAGTCGAGCAGCCCCAAACAGAGAGGATTTCACGCATCAATCA 4708
1517 uGlyLysAlaGlyLysGlnLeuCysMetPheArgThrCysSerCysGlyV 1534
|||||
4709 AGGCAAGGCGGTAAACAGCTTTGCATGTTTAGAACGTTCTCTCGGGTG 4758
1534 aGlnLeuAspValTyrAsnGluAlaThrIleAlaThrArgPheSerAsn 1550
|||||
4759 TGCAGCTGGATGTGTACAACGAGCGGACTATCGCCACCAAGGTCTCAAAAC 4808
1551 AlaPheThrPheValAspAsnLeuLysGlyArgSerAlaValPhePheSe 1567
|||||
4809 GCATTTACCTTTGTCGATAACTTGAAGGAGGAGGAGTGGCGTCTTTTCTC 4858
1567 rLysLeuGlyGluGlyTyrThrTyrAsnGlyGlySerHisValSerSerG 1584
|||||
4859 AAAGTGGGTGAGGGGTATACCTATATAATGGTGGTAGCCATGTTTCATCAG 4908
1584 lyTrpProArgAlaLeuGluAspIleLeuThrAlaIleLysTyrProSer 1600
|||||
4909 GGTGGCCCTCGTCCCTAGAGGATATCTTAACGGCAATTAAGTACCCAGC 4958
1601 ValPheAspHisCysLeuValGlnLysTyrLysMetGlyGlyValPr 1617
|||||
4959 GTCCTTCGACCACTGTTTAGTCAGAGAGTACAAGATGGGTGGAGCGGTACC 5008
1617 oPheHisAlaAspAspGluGluCysTyrProSerAspAsnProIleLeuT 1634
|||||
5009 ATTCCAGCGCTGATCAGGAGGAGTGCTATCCATCAGATAACCCCTATCTTCA 5058
1634 hrValAsnLeuValGlyLysAlaAsnPheSerThrLysCysArgLysGly 1650
|||||

5059 CGGTCAATCTCGTGGGGAAGGCAAACTTCTCGACTAAGTCACGAAGGGT 5108
1651 GlyLysValMetValIleAsnValAlaSerGlyAspTyrPheLeuMetPr 1667
|||||
5109 GGTAAAGTCATGGTCATAAACGTAGCTTCGGGTGACTATTTCTTATGCC 5158
1667 oCysGlyPheGlnArgThrHisLeuHisSerValAsnSerIleAspGluG 1684
|||||
5159 TTGCGGTTTTCAAAGGACGCACTTGCAATCAGTAAACTCCATCGACGAAG 5208
1684 lyArgIleSerLeuThrPheArgAlaThrArgValPheGlyValGly 1700
|||||
5209 GGGCATACAGTTTGACGTTTCAGGGCAACTCGGGCGCTCTTTGGTGTAGGC 5258
1701 ArgMetLeuGlnLeuAlaGlyGlyValSerAspGluLysSerProGlyVa 1717
|||||
5259 AGGATGTCAGTTACCGCGCGGTGTCGGATGAGAAGTCACCAAGTGT 5308
1717 lProAsnGlnGlnProGlnSerGlnGlyAlaThrArgThrIleThrProL 1734
|||||
5309 TCCAAACCAACCAACCAACAGAGCAAGGTGCTACCAACAATCACACAA 5358
1734 ySSerGlyGlyLysAlaLeuSerGluGlySerGlyArgGluValLysGly 1750
|||||
5359 AATCGGGGGGCAAGGCTCTATCTGAGGGGAAGTGGTAGGGAAAGTCAAAGGG 5408
1751 ArgSerThrTyrSerIleTyrCysGluGlnAspTyrValArgLysCysG 1767
|||||
5409 AGGTCACATACCTCGATATCGTCCGAACAAGATTACGTTAGGAAGTCTGA 5458
1767 uTrpLeuArgAlaAspAsnProValMetAlaLeuGluProAspTyrThrP 1784
|||||
5459 GTGGCTCAGGGCTGATAATCCAGTGATGGCTCTTGAACCTGACTACACCC 5508
1784 rMetThrPheGluValValLysThrGlyThrSerGluAspAlaValVal 1800
|||||
5509 CAATGACATTTGAAGTGGTTAAACCGGGACCTCTGAAGATCCGCTCGTG 5558
1801 GluTyrLeuLysTyrLeuAlaIleGlyIleGluArgThrTyrArgAlaLe 1817
|||||
5559 GAGTACTTCAAGTATCTGGCTATAGCATTTAGAGGACATACAGCGGCTT 5608
1817 uLeuMetAlaArgAsnIleAlaValThrThrAlaGluGlyValLeuLysV 1834
|||||
5609 GCTTATGGCTAGAAATATTGCCGCTACTACCGCCGAAGGTGTTCTGAAAG 5658
1834 aProAsnGlnValTyrGluSerLeuProGlyPheHisValTyrLysSer 1850
|||||
5659 TACCTAATCAAGTTTATGAATCACTACCGGGCTTTCACGTTTACAAGTCG 5708
1851 GlyThrAspLeuIlePheHisSerThrGlnAspGlyLeuArgValArgAs 1867
|||||
5709 GGCACACATCTCATTTTTCATTCACACAAAGACGGCTTCGCTGTGAGAGA 5758
1867 pleuProTyrValLeuIleAlaGluLysGlyIlePheThrLysGlyLysA 1884
|||||
5759 CCTACCGTAGCTACTCATAGCTGAAAAGGTATCTTTTACCAGGGGCAAG 5808
1884 spValAspAlaValAlaLeuGlyAspAsnLeuPheValCysAspAsp 1900
|||||
5809 ATGTCACGCGGTGGTAGCTTTGGGCGACAATCTGTTCGTATTCGACGAT 5858
1901 IleLeuValPheHisAspAlaIleAsnLeuIleGlyAlaLeuLysValAl 1917
|||||
5859 ATACTGGTTTTCCACCATGCCATTAAATTTGATAGGTGCACCTGAAAGTCCG 5908
1917 aArgCysGlyMetValGlyCysSerPheLysSerPheGluTyrLysCyst 1934
|||||
5909 TCGATCGGCATGGTGGCGAATCGTTTTAAGTCTCTCGAATATAAGTGCT 5958
1934 yAsnAlaProProGlyGlyLysThrThrThrLeuValAspGluPhe 1950
|||||

```
5959 ATAATGCTCCCCAGGTGGCGGTAAAGACGACGACGCTTAGTGGACGAATTC 6008
1951 ValLysSerProAsnSerThrAlaThrIleThrAlaAsnValGlySerSe 1967
|||||
6009 CTTAAGTCAACCAATAGCACAGCCACCAATTACGGCTAATGTGGGAAGTTC 6058
1967 rGluAspIleAsnMetAlaValLysLysArgAspProAsnLeuGluGlyL 1984
|||||
6059 TCAGGCACATAAATATGCGCGTGAAGAAGAGATCCGAATTTGGAAGGTC 6108
1984 euAsnSerAlaThrThrValAsnSerArgValValAsnPheIleValArg 2000
6109 TCAACAGTGCACACAGATTAACTCCAGGGTGGTAAACTTTATCGTCAGG 6158
2001 GlyMetTyrLysArgValLeuValAspGluValHisMetMetHisGlnG1 2017
6159 GCAATGCTATAAAGCGTTTGGTGGATGAGGTGCACATGATGCATCAAGG 6208
2017 yLeuLeuGlnLeuGlyValPheAlaThrGlyAlaSerGluGlyLeuPheP 2034
|||||
6209 CTTACTACAACCTAGGCGTCTTCGCAACCGCGCGTCCGAAGCGCTCTTTT 6258
2034 heGlyAspIleAsnGlnIleProPheIleAsnArgGluLysValPheArg 2050
6259 TTGGAGACATAAATCAGATACCATTCATAAACAGGGAGAGGTGTTTAGG 6308
2051 MetAspCysAlaValPheValProLysLysGluSerValValTyrThrSe 2067
6309 ATGGATTCTGCTGTTTGTTCCTCAAGAGGAAGCGTTGTATACACTTC 6358
2067 rLysSerTyrArgCysProLeuAspValCysTyrLeuLeuSerSerMetT 2084
6359 TAAATCGTACAGGTCTCCGCTTAGATGTTTGTCTACTTGTTCCTCAATGA 6408
2084 hrValArgGlyThrGluLysCysTyrProGluLysValValSerGlyLys 2100
6409 CCGTAAGGGGACCGGAAAGTGTACCCGTGAAAGGTCGTTAGCGGTAAAG 6458
2101 AspLysProValValArgSerLeuSerLysArgProIleGlyThrThrAs 2117
6459 GACAAACCACTAGTAAAGATCCCTGTCCAAAAGGCCAATTTGGAACCACTGA 6508
2117 pAspValAlaGluIleAsnAlaAspValTyrLeuCysMetThrGlnLeug 2134
6509 TGACCTAGCTCAAAATAAACCGCTGACGTCTACTTGTGCATGACCCAGTTGG 6558
2134 luLysSerAspMetLysArgSerLeuLysGlyLysGlyLysGluThrPro 2150
6559 AGAAGTCGGATATGAAGAGGTGCTTGAAGGGGAAAGGAAAGAAACACCA 6608
2151 ValMetThrValHisGluAlaGlnGlyLysThrPheSerAspValValLe 2167
6609 GTGATGACAGTGCATGAAGACACAGGAAACATTCAGTGATGTGTATT 6658
2167 uPheArgThrLysLysAlaAspSerLeuPheThrLysGlnProHisI 2184
6659 GTTTAGGACGAAGAAAGCCGATGACTCCCTATTCTACTAAACAACCGCATA 6708
2184 leLeuValIcLysSerArgHisThrArgSerLeuValTyrAlaAlaLeu 2200
6709 TACTTGTGTTGTTGTCGAGACACACACCGCTCACTGGTTTATGCCGCTCTG 6758
2201 SerSerLysLeuAspAspLysValGlyThrTyrIleSerAspAlaSerPr 2217
6759 AGCTCAAAAGTTGGACGATAAGGTCCGACATATATATTACGACGCGTCAAC 6808
2217 oGlnSerValSerAspAlaLeuLeuHisThrPheAlaProAlaGlyCysP 2234
6809 TCAATACGATATCCGACGCTTGTCTTCACACGTTCCGCCCGCGCTGTTGCT 6858
2234 hcArgGlyIle 2237
|||||
6859 TTCGAGGTATA 6869
```

```
seq_name: /cgnl_8/gcgdata/geneseq/geneseq/NA1997.DAT.T72214
seq_documentation_block:
ID_T72214 standard; cDNA; 4173 BP.
XX T72214;
XX AC
XX XX
DT 19-SEP-1997 (first entry)
XX
DE Grapevine leafroll virus helicase cDNA.
XX
KW GLRaV; grapevine; Vitis; rootstock; leafroll; disease resistance;
KW transgenic plant; tristeza virus; citrus; helicase; ds.
XX
OS Grapevine leafroll associated virus type 3 isolate NYL.
XX
PN W09722700-A2.
XX
XX 26-JUN-1997.
XX
XX 20-DEC-1996; 96WO-US20747.
XX
PR 21-DEC-1995; 95US-0009008.
XX
XX (CORR ) CORNELL RES FOUND INC.
XX
XX Gonsalves D, Ling K;
XX
XX WPI; 1997-341691/31.
XX
XX P-PSDB; W21636.
XX
XX DNA encoding grape-vine leaf-roll virus proteins - useful to impart
XX viral-resistance to Vitis scion or root-stock cultivar(s)
XX
XX Claim 5; Page 66-69; 172pp; English.
XX
XX A double-stranded cDNA molecule (T72214) comprises the incomplete
XX coding sequence of the helicase (W21636) of grapevine leafroll
XX associated virus type 3 (GLRaV-3). It was identified as open
XX reading frame in a 15,227-nucleotide sequence that covers about
XX 80% of the GLRaV-3 genome. This sequence was isolated from a cDNA
XX library prep. from GLRaV-3 dsRNA obt. from leafroll-diseased
XX grapevine canes. Isolated GLRaV-3 nucleic acids (T72214-25) can be
XX used to prepare GLRaV-3 polypeptides (W21636-47), to impart GLRaV
XX resistance to Vitis scion or rootstock cultivars or tristeza virus
XX resistance to citrus scion or rootstock cultivars, and to design
XX probes for detection of GLRaV.
XX
XX Sequence 4173 BP; 1030 A; 874 C; 1177 G; 1092 T; 0 other;
XX
alignment_scores:
Quality: 5247.00 Length: 1423
Ratio: 4.214 Gaps: 17
Percent Similarity: 87.491 Percent Identity: 74.139
alignment_block:
US-09-301-906-15 x T72214 ..
Align seg 1/1 to: T72214 from: 1 to: 4173
848 ValSerValTyrAlaLysAlaMetValHisAspAsnPheAsnValLeuG1 864
|||||
1 GTGCTACTTTACGGCAAGAGTGTGATGAACGACAAATTTCAATATATCTTGA 50
864 uThrLeuMetSerMetProArgAlaPheIleArgLysValProGlySerV 881
|||||
51 GACCTGTGTAACCTTTGCCCAAGTCCTTTATAGTCAAGTACTGTTGTCGG 100
881 alValValThrIleCysThrSerGlyAlaSerAspArgLeuLeuArg 897
|||||
101 TGCTGGTTAGCATACCACTTCGGGCATTTCCGACAAACTTGAACCTCGG 150
```



```

2081 rSerMetThrValArgGlyThrGluLysCysTyrProGluLysValValS 2098
|||||
3702 CTCAATGACCGTAAAGGGAACGAAAGTGTACCCCTGAAAGGTCGTTA 3751
|||||
2098 erGlyLysAspLysProValValArgSerLeuSerLysArgProIleGly 2114
|||||
3752 GCGGTAAAGCAACACCACTAGTAGTACGCTGTCCAAAGGCCCAATTGGA 3801
|||||
2115 ThrAspAspValAlaGluIleAsnAlaAspValTyrLeuCysMetTh 2131
|||||
3802 ACCACTGATCAGCTAGCTGCTAAATAAACGCTGACGTGTACTTGTGCATGAC 3851
|||||
2131 rGlnLeuGluLysSerAspMetLysArgSerLeuLysGlyLysGlySG 2148
|||||
3852 CCAGTTGGGAGAGTCGGATATGAAGAGGTGCTTGAAGGGGAAAGGAAAG 3901
|||||
2148 luThrProValMetThrValHisGluAlaGlnGlyLysThrPheSerAsp 2164
|||||
3902 AAACACGATGATGACAGTGCATGAAGCACACAGGGAACATTCAGTGAT 3951
|||||
2165 valValLeuPheArgThrLysLysAlaAspAspSerLeuPheThrLysG1 2181
|||||
3952 GTGGTATTGTTTAGGACGAAGAAGCCGATGACTCCCTATTCTACTAAACA 4001
|||||
2181 nProHisIleLeuValGlyLeuSerArgHisThrArgSerLeuValTyrA 2198
|||||
4002 ACCGCATATACTTGTGGTTGTCGAGACACACACGCTCACTGGTTATG 4051
|||||
2198 laAlaLeuSerSerLysLeuAspAspLysValGlyThrTyrIleSerAsp 2214
|||||
4052 CCCTCTGACCTCAGATGGACCATTAAGTCGGCACATATATTACGGAC 4101
|||||
2215 AlaSerProGlnSerValSerAspAlaLeuLeuHisThrPheAlaProAl 2231
|||||
4102 GCCTCGCTCAATCAGTATCCGACGCTTGTTCACACAGTTCGCCCGCGC 4151
|||||
2231 aGlyCysPheArgGlyIle 2237
|||||
4152 TGGTCTCTTCGAGGTATA 4170

```

seq_name: /cgnl_8/gcgdata/geneseq/geneseq/NA2000.DAT:249205

seq_documentation_block:

ID 249205 standard; DNA; 873 BP.

AC XX

XX

DT XX

07-MAR-2000 (first entry)

DE XX

GLRaV-3 polyprotein helicase domain DNA.

KW XX

Grapevine; leafroll; grapevine leafroll-associated virus 3; GLRaV-3;

KW XX

viral disease; yield loss; sugar content; inhibition; infection;

KW XX

replication; polyprotein; domain; proteinase; methyltransferase;

KW XX

helicase; RNA-dependent; RNA polymerase; untranslacted region;

KW XX

transgenic plant; component; resistant; truncation; deletion; antisense;

KW XX

expression; detection; antibody; ds.

OS XX

Grapevine leafroll-associated virus 3.

PH XX

Key Location/Qualifiers

FT mat_peptide

1..873

FT /*tag= a

FT /product= "GLRaV-3 polyprotein helicase domain"

XX

WO9955880-A1.

XX

04-NOV-1999.

XX

29-APR-1999; 99WO-US09307.

XX

29-APR-1998; 98US-0083404.

PN

29-APR-1998; 98US-0083404.

PD

29-APR-1998; 98US-0083404.

PF

29-APR-1998; 98US-0083404.

PR

29-APR-1998; 98US-0083404.

XX (CORR) CORNELL RES FOUND INC.

XX

PI Gonsalves D, Ling K;

XX

WPI: 2000-062035/05.

DR

P-PSDB; Y58145.

XX

Newly isolated grapevine leafroll virus protein or polypeptide useful for producing transgenic plants conferring viral disease resistance -

XX

Claim 16; Fig 10; 84pp; English.

XX

This sequence represents DNA encoding the helicase domain of the polyprotein (Y58148) from the grapevine leafroll-associated virus 3 (GLRaV-3). Leafroll is a serious viral disease, occurring wherever grapes are grown. Although not lethal, it causes yield losses and reduction in sugar content. The virus encodes several proteins, which may serve as targets for the inhibition of viral infection or replication. These proteins include the 242-248 kD polyprotein (Y58148), encoded by open reading frame (ORF) 1a, and which comprises a proteinase domain (Y58143), a methyltransferase domain (Y58144), and a helicase domain (Y58145); a RNA-dependent RNA polymerase (Y58146), encoded by ORF 1b; and a protein of unspecified function (Y58147), encoded by ORF 11. Nucleotides encoding these proteins, or fragments thereof, and the 5' and 3' untranslated regions (UTRs) of the genome (249201-249202) are useful for the generation of transgenic plants and plant components. Such transgenic plants may be resistant to viral disease, for example, this property being conferred on the plants via the use of nucleotides encoding truncated or internally deleted proteins, or via the use of antisense nucleotides to inhibit viral gene expression. The nucleotides may additionally be used in the detection of viral nucleic acids in a tissue sample. The proteins can be used to generate antibodies which may be used to detect GLRaV-3 in plant samples. The isolation of GLRaV-3 facilitates the production of agents that reduce the risk of infection or damage by the virus in vineyards.

XX Sequence 873 BP; 262 A; 163 C; 233 G; 215 T; 0 other;

alignment_scores:

Quality: 1495.00

Length: 291

Ratio: 5.137

Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-301-906-15 x 249205

Align seg 1/1 to: 249205 from: 1 to: 873

1922 ValGlyGluSerPheLysSerPheGluTyrLysCysTyrAsnAlaProPr 1938

|||||

1 GTGGCGGAATCGTTAAGTCCTTCGAATATAAGTGTATATGCTCCCCC 50

1938 oGlyGlyGlyLysThrThrThrLeuValAspGluPheValLysSerProA 1955

|||||

51 AGTGGCGGTAGACGACGACGTTAGTGACGAATTCGTTAAGTCACCCA 100

1955 snSerThrAlaThrIleThrAlaAsnValGlySerSerGluAspIleAsn 1971

|||||

101 ATAGCACACGCCACCATTTACGGCTAATGTGGGAAGTCTGAGGACATAAAT 150

1972 MetAlaValLysLysArgAspProAsnLeuGluGlyLeuAsnSerAlaTh 1988

|||||

151 ATGGCGGTGAAGAAGAGAGATCCGAATTTGGAAGGCTCAACAGTGTCTAC 200

1988 rThrValAsnSerArgValValValAsnPheIleValArgGlyMetTyrLysA 2005

|||||

201 CACAGTTAACTCCAGGCTGCTAAACCTTTATCGTCAGGGGAATGTATAAAA 250

2005 rgValLeuValAspGluValHisMetMetHisGlnGlyLeuLeuGlnLeu 2021

|||||

251 GGGTTTTGGTGGATGAGGTGCACATGATGATCAAGGCTTTACTACAACTA 300

```

2022 GlyValPheAlaThrGlyAlaSerGluGlyLeuPhePheGlyAspIleAs 2038
|||||
301 GCGCTCTTCGCAACCGCGCGTCGGAAGCCCTCTTTTGGAGACATAAA 350
|||||
2038 nGlnIleProPheIleAsnArgGluLysValPheArgMetAspCysAlaVal 2055
|||||
351 TCAGATACCATTCATAAACAGAGGGAAGGTGTTAGGATGGATTGCTG 400
|||||
2055 aPheValProLysLysGluSerValValThrSerLysSerTyrArg 2071
|||||
401 TTTTGTTCCTCCAAAGAGGAAAGCGTTGTATACACTTCTAAATCGTACAGG 450
|||||
2072 CysProLeuAspValCysTyrLeuLeuSerSerMetThrValArgGlyTh 2088
|||||
451 TGTCCGTTAGATGTTTGTCTACTTGTCTCCTCAATCACCGTAAGGGGAAC 500
|||||
2088 rGluLysCysTyrProGluLysValValSerGlyLysAspLysProValV 2105
|||||
501 GGAAGAAGTGTACCTGAAAGGTCGTTAGCGGTAAGGACAAACCCAGTAG 550
|||||
2105 aArgSerLeuSerLysArgProIleGlyThrThrAspValAlaGlu 2121
|||||
551 TAAGATCGCTGTCCAAAGGCCAATTGAACCACTGATGACGTAGCTGAA 600
|||||
2122 IleAsnAlaAspValTyrLeuCysMetThrGlnLeuGluLysSerAspMe 2138
|||||
601 ATAAACGCTGACCTGCTACTTGTGCATGCCACCTTGGAGAGTCGGATAT 650
|||||
2138 tLysArgSerLeuLysGlyLysGlyLysGluThrProValMetThrValH 2155
|||||
651 GAAGAGGTGCTTGAAGGAAAGGAAAGAAACACCACTGATGACAGTGC 700
|||||
2155 iGluAlaGlnGlyLysThrPheSerAspValValLeuPheArgThrLys 2171
|||||
701 ATGACGACAGGGAACAAACATTCAGTGATGGTATGTTTAGGACGAG 750
|||||
2172 LysAlaAspAspSerLeuPheThrLysGlnProHisIleLeuValGlyLe 2188
|||||
751 AAAGCCGATGACTCCCTATTCACTAAACACCGCATATCTATTGTTGGTT 800
|||||
2188 uSerArgHisThrArgSerLeuValTyrAlaAlaLeuSerSerLysLeuA 2205
|||||
801 GTCGAGACACACACGCTCACTGGTTTATGCCGCTCTGAGCTCAAAAGTTGG 850
|||||
2205 sAspLysValGlyThrTyrIle 2212
|||||
851 ACGATAAGGTCGGCACATATATT 873

```

seq_name: /cgnl_8/gcgdata/geneseq/geneseqn/NA2000.DAT:Z49204

seq_documentation_block:

ID Z49204 standard; DNA; 816 BP.

XX

AC Z49204;

XX

07-MAR-2000 (first entry)

XX

GLRav-3 polyprotein methyltransferase domain DNA.

XX

KW Grapevine; leafroll; grapevine leafroll-associated virus 3; GLRav-3;
KW viral disease; yield loss; sugar content; inhibition; infection;
KW replication; polyprotein; domain; proteinase; methyltransferase;
KW helicase; RNA-dependent; RNA polymerase; untranslated region;
KW transgenic plant; component; resistant; truncation; deletion; antisense;
KW expression; detection; antibody; ds.

XX

OS Grapevine leafroll-associated virus 3.

XX

Key Location/Qualifiers

PH mat_peptide 1..816

FT /*tag= a

FT /product= "GLRav-3 polyprotein methyltransferase domain"

XX

PN WO9955880-A1.

XX

PD 04-NOV-1999.

XX

PF 29-APR-1999; 99WO-US09307.

XX

PR 29-APR-1998; 98US-0083404.

XX

PA (CORR) CORNELL RES FOUND INC.

XX

PI Gonsalves D, Ling K;

XX

DR WPI; 2000-062035/05.

XX

PS P-PSDB; Y58144.

XX

PT Newly isolated grapevine leafroll virus protein or polypeptide useful
PT for producing transgenic plants conferring viral disease resistance -
PS Claim 14; Fig 7; 84pp; English.

XX

CC This sequence represents DNA encoding the methyltransferase domain of the
CC polyprotein (Y58148) from the grapevine leafroll-associated virus 3
CC (GLRav-3). Leafroll is a serious viral disease, occurring wherever
CC grapes are grown. Although not lethal, it causes yield losses and
CC reduction in sugar content. The virus encodes several proteins,
CC which may serve as targets for the inhibition of viral infection or
CC replication. These proteins include the 242-248 kD polyprotein (Y58148),
CC encoded by open reading frame (ORF) 1a, and which comprises a proteinase
CC domain (Y58143), a methyltransferase domain (Y58144) and a helicase
CC domain (Y58145); a RNA-dependent RNA polymerase (Y58146), encoded by ORF
CC 1b; and a protein of unspecified function (Y58147), encoded by ORF 11.
CC Nucleotides encoding these proteins, or fragments thereof, and the 5',
CC and 3' untranslated regions (UTRs) of the genome (Z49201-Z49202) are
CC useful for the generation of transgenic plants and plant components.
CC Such transgenic plants may be resistant to viral disease, for example,
CC this property being conferred on the plants via the use of nucleotides
CC encoding truncated or internally deleted proteins, or via the use of
CC antisense nucleotides to inhibit viral gene expression. The nucleotides
CC may additionally be used in the detection of viral nucleic acids in a
CC tissue sample. The proteins can be used to generate antibodies which
CC may be used to detect GLRav-3 in plant samples. The isolation of
CC GLRav-3 facilitates the production of agents that reduce the risk of
CC infection or damage by the virus in vineyards.

XX

SQ Sequence 816 BP; 183 A; 162 C; 234 G; 237 T; 0 other;

alignment_scores:

Quality: 1432.00 Length: 272

Ratio: 5.265 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-301-906-15 x Z49204 ..

Align seg 1/1 to: Z49204 from: 1 to: 816

460 LeuLysProArgGluArgGluLysLeuArgGluLeuPheProGluLeuSe 476

|||||

1 CTGAAGCCGCGGGAAGGAGAGCTGAGGGAACCTTTCCAGAGCTTTC 50

|||||

476 rIleGlnPheSerAspSerValArgSerHisProPheAlaAsnAlaM 493

|||||

51 GATACAGTCTCCGACTCGGTGAGGAGTAGTCACCCATTCGCTAATGCCA 100

|||||

493 etArgSerCysPheAsnGlyIlePheSerArgArgCysGlyAsnValCys 509

|||||

101 TCGCGAGCTGTTTCAATGGAATCTTTTCAGGAGGTGGTAAATGTGTCC 150

|||||

510 PhePheAspIleGlySerPheThrThrHisValLysAlaGlyHisVa 526

|||||

151 TTCTTCGATATTGGGGGAGCTTCACGTATCATGTCAAGCTGGCCATGT 200

526 lAsnCysHisValCysAsnProValLeuAspValLysAspValLysArgA 543
 201 GAACCTGTCATGATATGCAATCCAGCTCCCTAGACGTTAAAGATGTAAGCGGA 250
 543 rGleAsnGluIleLeuPheLeuSerThrAlaGlyGlyAspSerTyrVal 559
 251 GAATCAATGAGATCCTCTTTCTTCCACAGCTGGGGAGATTCGTACGTG 300
 560 SerSerAspLeuLeuThrGluAlaAlaSerLysSerValSerTyrCys 576
 301 TCCAGTGCACCTTCTAACTGAAGCGGCTTCAAAGCTCTGCTTACTGTAG 350
 576 rArgLysSerGlnAsnCysAspSerArgAlaAspAlaGlyPheMetVal 593
 351 TCGAGAAATCCAGAACTGCGATTCTAGAGCCGATGCGGGTTTATGGTGG 400
 593 spValTyrAspIleSerProGlnGlnValAlaGluAlaMetAspLysLys 609
 401 ATGTGTACGATATATCCCCGACAGGTACGAGAGCTATGGATAAGAAG 450
 610 GlyAlaLeuValPheAspIleAlaLeuMetPheProValGluLeuLys 626
 451 GCTCGCGTGGTTTTTCGACATAGCTCTTATGTTCCCGTGGAGTTGTTGA 500
 626 rGlyAsnGlyGluValTyrLeuGluGluLeuAspThrLeuValLysArgG 643
 501 CGGTAAACGGGTCAAGTTTACTTTGGAAGAACTCGATACGTTGGTGAAGAGG 550
 643 luGlyAspTyrLeuAlaTyrAsnValGlyGlnCysGlyGluMetTyrGlu 659
 551 AAGGTGATTACCTGGCCTACAATGTTGGTCAGTCTGTCAGATGATGAA 600
 660 HisSerPheSerAsnValSerGlyPhePheThrPheSerTyrValArg 676
 601 CATTCCTCTCTAACGTAAGCGGTTTTTCCACCTTTTCTATGTACGCAC 650
 676 rSerSerClyAsnValPheLysLeuGluTyrGluGlyTyrArgCysGlyT 693
 651 TTCGTCGGGAACGTGTTTAAAGTAGAGTATGAGGATACCGTGTGGTT 700
 693 yrHisHisLeuThrMetCysArgAlaGlnLysSerProGlyThrGluVal 709
 701 ACCATCATCTCACTATGTGTAGGCTCAGAAGTCACCTGGAACCTGAGTT 750
 710 ThrTyrArgSerLeuValProSerPheValGlyLysSerLeuValPheI 726
 751 ACGTATAGGTGTTGGTCCGCTGCTTCGTGGGCAAAATCGCTGGTGTTCAT 800
 726 eProValValAlaGly 731
 801 ACCTGTTGTAGCTGT 816

seq_name: /cgnl_8/gcgdata/geneseq/geneseqn/NA1999.DAT:V08864

seq_documentation_block:

ID V08864 standard; cDNA; 7920 BP.

XX AC V08864;

XX DT 29-MAR-1999 (first entry)

XX DE Grapevine leafroll virus type 2 (GRLAV-2) ORF1a.

XX KW GRLAV-2; closterovirus; grape; tobacco; transgenic plant; disease resistance; virus resistance; beet yellows virus; tristeza virus; protease; methyltransferase; helicase; ss.

XX OS Grapevine leafroll virus type 2.

XX PN W09853055-A1.

XX PD 26-NOV-1998.

793	SerProAspAspMetTrpGlyLeuValValAlaValMetAlaGlnAlaIle	809
794	
795	
796	
797	
798	
799	
800	
801	
802	
803	
804	
805	
806	
807	
808	
809	
810	
811	
812	
813	
814	
815	
816	
817	
818	
819	
820	
821	
822	
823	
824	
825	
826	
827	
828	
829	
830	
831	
832	
833	
834	
835	
836	
837	
838	
839	
840	
841	
842	
843	
844	
845	
846	
847	
848	
849	
850	
851	
852	
853	
854	
855	
856	
857	
858	
859	
860	
861	
862	
863	
864	
865	
866	
867	
868	
869	
870	
871	
872	
873	
874	
875	
876	
877	
878	
879	
880	
881	
882	
883	
884	
885	
886	
887	
888	
889	
890	
891	
892	
893	
894	
895	
896	
897	
898	
899	
900	
901	
902	
903	
904	
905	
906	
907	
908	
909	
910	
911	
912	
913	
914	
915	
916	
917	
918	
919	
920	
921	
922	
923	
924	
925	
926	
927	
928	
929	
930	
931	
932	
933	
934	
935	
936	
937	
938	
939	
940	
941	
942	
943	
944	
945	
946	
947	
948	
949	
950	
951	
952	
953	
954	
955	
956	
957	
958	
959	
960	
961	
962	
963	
964	
965	
966	
967	
968	
969	
970	
971	
972	
973	
974	
975	
976	
977	

1054GluValSerLeuAsnAs 1059
3571 AGTCGTGTCGCTAGCGCGGTGTGCTGCTGGAAGAAGTTGCTACCTCATG 3620
1059 pPheThrThrPheValLeuArgAsnLysValLeuIleGlyIlePheValA 1076
.....III:::IIIIII:::III
3621 CGCGAAGCGGAGCGTTCTTAAGACCAAGGTATGATTT..... 3657
1076 laSerLeuGlyAlaAlaProIleAlaThrLysTyArgArg..GlyIle 1091
:::IIIIII:::III:::IIIIII:::III
3658 ..ACCTTCGACGCGCGCTGTGCTATGATGTTTAAATAGCTCGCGTTT 3705
1092 AlaAlaAsnAlaArgTyAlaGly..... 1100
:::IIIIII:::III:::IIIIII:::III
3706 TCAGCGACGCTCGGAGTATAAAATCGTATATACATCGTTACACGCAAGT 3755
1101SerSerTyGluThrLeuSerS 1108
IIIIIIIIII:::IIIIII:::IIIIII:::III
3756 ATTCTTTTGACACTATCTTTTTTTGAGGACAGCAGTTACCTACCCATAGAAG 3805
1108 erLeuSerSerGlnAlaAlaGlyGlyLeuArgGlyLeuThrSerSer... 1123
IIIIIIIIII:::IIIIII:::III:::IIIIII:::III
3806 TTCGAGTTCGGCGATATGCGGTGTCATCGTCACACTTTTCCTCGGCG 3855
1124ThrValSerGlyGlySerLeuValValArgArgGlyPheS 1137
:::IIIIII:::III:::IIIIII:::III
3856 TCGTCCATAAGTTTAAACGCTTCTTACTTCAAATTAACCAAGGATCTC 3905
1137 rSerAlaValThrValThrArgAlaThrValAlaLysArgGlnValProL 1154
IIIIIIIIII:::IIIIII:::IIIIII:::III
3906 CCTAGAGGTTCGCTC.....CGAATGTTTGTGC 3934
1154 euAlaLeuLeuSerPheSerThrSerTyAlaIleSerGlyCysSerMet 1170
:::IIIIII:::III:::IIIIII:::III
3935 GAGTCACGCATGGTTTGACACCA...GCACCGACGCGCTC...ATA 3978
1171 LeuGlyIleTrpAlaHisAlaLeuProArgHisLeuMet..... 1183
IIIIIIIIII:::IIIIII:::IIIIII:::III
3979 CGTGGGTTTTCTCCCA...ATTGTGCTCACTTACTTGTGGAAATAC 4025
1184PhePhePheGlyLeuGlyThrLeuL 1192
IIIIII:::IIIIII:::IIIIII:::III
4026 GGTAAATGTGGCTTTACCACTCAGCTTTCATCCCGGGTGGTGCCTTT 4075
1192 euGlyAlaArgAlaSerAlaAsnThrTrp.....LysPhe 1203
IIIIII:::IIIIII:::IIIIII:::III
4076 TAGTTAAAGTGTGTGAGCTTAATCTTCATCTTCGGTGAAGATCTTAT 4125
1204 GlyGlyPheSerAsnAsnTrpCysAlaValProGluValValTrp.... 1218
:::IIIIII:::III:::IIIIII:::III
4126 TCCGGTTTTATTAAAGCAC....GGAATCAGTGAATCTCTTTCTCTTAG 4169
1219ArgGlyLysSerValSerSerLeuLeuLeuP 1229
:::IIIIII:::IIIIII:::IIIIII:::III
4170 TAGTATTCTGAAGTCTTGAAGGGTAAGCTTGTGACGAGTTG..... 4212
1229 rolLeThrLeuGlyValSerLeuIleIleArgGlyLeu..... 1241
:::IIIIII:::IIIIII:::IIIIII:::III
4213AAATCGATTATTCAGGGGTTTTTGAATCCCAAC 4245
1242LeuAsnAspThrIlePr 1247
:::IIIIII:::IIIIII:::IIIIII:::III
4246 AAGCAGCTGTTTAAAGAAGCTACTCAGGAAGCGATTTCGTCACGCGTAT 4295
1247 oGln.....LeuAlaTyValProValGlu.....GlyArgA 1259
IIII:::IIII:::IIII:::III
4296 GCAAGTGCCTGTCTGCTAGTGTGATGCGCTTAAGAGCGCGCGGGAATA 4345
1259 snValTyAspGluThrLeuArgTyTyArgAspPheAspTyArgGlu 1275
:::IIIIII:::IIIIII:::IIIIII:::III
4346 TTTATACAATTTTCTAGTCTGA.....CGTACCTTTGGTAGGATGAA 4389
1276 GlyAlaGlyProSerGlyThrGlnHisGluAlaValProGlyAspAspAs 1292

1744	rGlyArgGluValLysGlyArgSerThrTySerIleTrpCysGluGlnA	176
5997	CAATAGACGC...AGAAATCGTGCACACTAAT...	6031
1761	sptYrValArgLys.....Cys	1766
6032	AATTTACGGCCCTGCTCATGACGAGATGCAAGAAGTCACCACCTTTCTGTG	6081
1767	GluTrpLeuArgAlaAspAsnProValMetAlaLeuGlu.....	1779
6082	TCC.....AAAGCCCAACTCTGCTGGTTTGGCCCTTGGAAAGGCGAGTGCT	6125
1780ProAspTyThrPrometThrPheGluV	1789
6126	TGTGGAGAGCGTATAAAGTCGGGAAACTTCTTAAGACGGGTTAATGAGA	6175
1789	alValLysThrGly.....ThrSerGluAspAlaValVal.....	1800
6176	TGSGTAGGAAAGGGAGTACCACCGACGGAAGACTGGCGGTGCTTGTCTGCG	6225
1801GluTyrLeuLysTyThrLeuAlaIleGlyLeuGluArg.....	1812
6226	GACGATGACCGCTGGAAATCTCTGTTGCTCAGCGACGAGACGATTC	6275
1813ThrTyArg.....AlaL	1817
6276	GCCTAAGACAGCTCAGGATTAAGCGAATAGCTAAATAGGTTAAACTCAAGCT	6325
1817	euLeuMetAlaArgAsnIleAlaValThrAlaGluGlyValLeuLys	1833
6326	TGCAATTCCCGAAGCTTATTGTTGGACGACACAAGGATACCGGGGT	6375
1834	ValProAsnGlnValTyGluSerLeuProGlyPheIisValTyLysSe	1850
6376	CTAACCAACGCGCTGAGGAG.....TTTATTATATATGCAAGA	6413
1850	rGlyThrAspLeuIlePheHisSerThr.....GlnAspGlyL	1863
6414	ACTTGCTCTTTTGGAAATCCACGAACACTGTGCACCTACTACGTCAAC	6463
1863	euArgVal.....ArgAspLeu	1868
6464	TGCGCATAGTCAACTTCGATCGTTCGCTAGCACCATGCAGCGAAGATGCT	6513
1869	ProTyrrValLeuIleAlaGluLysGlyIlePheThrLysGlyLysAspVa	1885
6514	CAGCTGTACGTACGGAAGAAGCGGTCAACAGCATAGTCGAGGGTAAAGAGT	6563
1885	lAspAla...ValValAlaLeuGlyAspAsnLeuPheValCysAspAspI	1901
6564	ACGTTTGACATTAGGATTTCACGATTCACGATTTCTGTTGACGGAA	6613
1901	leLeu.....ValPheHisAsp	1906
6614	AAATTTCTATTAAACAGCGCGGCGAGCGCAAAATGTTTATATATACACG	6663
1907	AlaIleAsnLeuIleGlyAlaLeuLysValAlaArgCysGlyMetValGI	1923
6664AAACCTC.....CGCTTCTTGGCGAGTAAATTTGTTTATAGCCGG	6701
1923	yGluSerPhe.....LysSerPheGluTyrl	1932
6702	CTACCCCTTTTCAAGGAGCTTCGCTTTCACGAATTCGTGCGTTCGATATTC	6751
1932	ysCysTyAsnAlaProGlyGlyGlyLysThrThrThrLeuValAsp	1948
6752	TCTCTACGAAGCTCCACCGGAGGTGTGAAGACGACGACGCTGATTGAC	6801
1949	GluPheVal.....LysSerProAsnSerThrAlaThrIleTh	1961
6802	TCGTTCTTTGAGGTCTTCAAGAAGAGGTGAGGTTTTCACCATGATCTTAAAC	6851

1961	rAlaAsnValGlySerSerGluAspIleAsnMetAlaValLysLysArgA	1978
6852	CGCCAAAGAAATGGCGAGGTGAGATCCTAAAGAAAGTGGAGAACGAAG	6901
1978	spProAsnLeuGlu.....GlyLeuAsn	1985
6902	TGTCATAACATTGAATGCCAGAAAACGTAAAGACAAGAAGATCTCCGAAAAAG	6951
1986	SerAlaThrThrValAsnSerArgValValAsnPheIleValArgGlyMe	2002
6952	AGCATTTACACCACCATCGACGCTATTTAATGCATCAC.....CGTGCGTTG	6995
2002	tTyLysArgValLeu...ValAspGluValHisMetMetHisGlnClyL	2018
6996	TGATGCAGACGCTCTTTTCATCGATGAGGTTCATGGTTCATCGGGTA	7045
2018	euleuGlnLeuGlyValPheAlaThrGlyAlaSerGluGlyLeuPhePhe	2034
7046	GCGTACTAGCTTGCAATTCAGTTCACGAGGTGTCATAAAGCTAATGATCTTC	7095
2035	GlyaspIleAsnGlnIleProPheIleAsnArg.....GluLysVa	2048
7096	GGGATAGCCGCGAGATCACTACATTTGAAGGAACGAATTTGGACAAGTG	7145
2048	LpheArgMetAspCysAlaValPheValProLysLysGluSerValValt	2065
7146	TTTGATGGGATTCGACAGGTTTCGT...GACCTGCAGTTCGGGTTTT	7192
2065	yrrThrSerLysSerTyrrArgCysProLeuAspValCystyrLeuLeuSer	2081
7193	ATGGTAATATTTCGTACCGTTGTCATGGGATGTGTGCGCTTGGTTAAGC	7242
2082	SerMet.....ThrValArgGlyThrGluLysCysty	2092
7243	ACAGTGTATGGCAACCTAATCGCCACCGTGAAGGTGAAGCGCAA....	7287
2092	rProGluLysValValSerGlyLysAspLysProValValArgSerLeus	2109
7288	GGTAAGAGC.....AGCATGC	7303
2109	erLysArgProIleGlyThrThrAspAspValalaGlutIleAsnAlaAsp	2125
7304	GCATTACGAAATTAATTCAGTCGACGATTTAGTCCCACGCGTGGTTC	7353
2126	ValTyrrLeuCysMetThrGlnLeuGluLysSerAspMetLysArgSerLe	2142
7354	ACGPTTCTGTGTATGCTTCAGTCGGAGAAGTTGGAATCAGCAAGCACATT	7403
2142	uLysGlyLysGly...LysGluThrProValMetThrValHisGluAlaG	2158
7404	TATTCGCAAGGTTTGACTAAACTTAACGTTCTACGGTGCATGAGCGC	7453
2158	lnGlyLysThrPheSerAspValValLeuPheArgThrLysLysAlaAsp	2174
7454	AAGGTGAGACGATATCGCGGTGTGAACCTTGTGCGACTTAAGTTTCAGG	7503
2175	AspSerLeuPheThrLysGlnProHisIleLeuValGlyLeuSerArgHl	2191
7504	GATGAACCCCTTAATCATCATCAGGCACATAACCGTGCCTTCTTCGTCA	7553
2191	sThrArgSerLeuValTyrrAlaLeuSerSerLysLeuAspAspLysv	2208
7554	CACCGACAGCTTAACCTATACGCTTAGCTGCTCGTCGAGGTGACGCCA	7603
2208	alGlyThrTyrrIleSerAspAlaSer	2216
7604	CTTGCGATGCCATCCAGAAGGCTGCG	7629

```
seq_name: /cgn1_8/gcgdata/geneseq/NA1999_DAT.V08874
seq_documentation_block:
ID V08874 standard: cDNA; 15500 BP.
XX
```


1340 TGTTCCTTTCCGGCGGGCGCGTTCCTTTCCTTTA..... 1374
281 AsnThrAlaProGlyCysAspValAlaLeuProArgIleGluLeuTrpse 297
1375GTTTGTGACTAGATAGATAC..... 1398
297 rMetArgGluArgAlaPheValCysThrThrLysGlyTrpCysTrpPheA 314
1399CCGAACGGGTTTTTACTTGG 1420
314 snAsnGluArgLeuArgGlyGluIleArgArgArgCysPheSerSer 330
1421 CTCACCTCCGGTACCGCGTTCCTTAAGGGGTTTGTATCCGAAG 1470
331 SerPheSerIleGly.....Ph 336
1471 CGTTTCGACATCGGTCTTCCCCACCGCGCAAGCTCAGAAACCGTAT 1520
336 eLeuMetHisLeuGlyPheArgSerLeuLysValIleArgPheAlaGlyT 353
1521 GGTTCGCGAGCTTGGTGAAGAAAGTTTA..... 1548
353 hrAsnIleLeuHisMetProSerLeuAsnGluArgThrPheGlyTrp 369
1549GGTTTGAACCTGTACGGCGCATATACGTCA 1578
370 LysGly.....GlyAspValTy 375
1579 CGCGCGGTCTTTCACCTGCATATGACGCTAAGTTTATAAAGGATTTGCG 1628
375 rLeuProAsnValProLysThrAlaIleValAlaGlyAspArgThrArgL 392
1629 TCTT.....ATGTCAGCAGCTTATAGCTGGA...AAGCAGGGG 1663
392 euGlyGlyGluIleLeuAlaSerValAlaAsnAlaLeuAsnGlnGlu 408
1664 TGGAGAGGTGGTACCTCTTGACATACTCTGCCATGAAGCAGAA... 1710
409 ValTyrSerSerValValSerSerIleThrAsnArgLeuValLeuArgAs 425
1711ACGTCGAAGCGGTGTATGATAGATTATATGGCGGCAC 1748
425 pGlnSerAlaLeuLeuSerHisLeuAspThrLysLeuCysAspMetPheS 442
1749 TCACCTCGTTGCTGAAACTGACGATCAGCAGAAAGACTTAATCGATTCAAAA 1798
442 eGlnArgAspAlaMetIleArgGluLysProSerHisArgCysAspVal 458
1799 ATGACGTGCAGAGTTTGAAAGAGATCGGCGGATTCTCAAGTGCCTTT 1848
459 PheLeuLysProArgGluArgGluLysLeuArgGluLeuPheProGluLe 475
1849 TACATGTCGAAGCAACACAGAATTCGCTGACGCGTTTCTACCTCAGTT 1898
475 uSerIleGlnPheSerAspSerValArgSerSerHisProPheAlaAsnA 492
1899 CGAACTTAAGTTTTCACCTCGCATTCAGATCATCCCGCGCGCGC 1948
492 laMetArgSerCysPheAsnGlyIlePheSerArgArgCysGlyAsnVal 508
1949 CTCTAGACTGCTGGAATAATGAACCTTATGTGGCTTATGTGGTAAT... 1995
509 CysPhePheAspIleGlyGlySerPheThrThrHisValLysAla...GI 524
1996 AGCGTTTCAGATATCGAGGTGTCTCTCTTTCCATTTCGATTCACAGAC 2045
524 yHisValAsnCysHisValCysAsnProValLeuAspValLysAspVal 541
2046 GCAAGACGGGTTCACGTATGAGGCTGTGTGGATGCAAGCATCGCG 2095
541 ysArgArg...IleAsnGluIleLeuPheLeuSerThrAlaGlyGlyAsp 556
:||||| :||||| :|

2096 AGCGTCGCGTGGTGGTGATTTGCAGTATTTCCAACGTGCGTTGGGAGAC 2145
557 SerTyrValSerSerAspLeuLeuThrGluAlaAlaSerLysSerValse 573
2146GATGATAAAATTTTGGAA...GGGCCACCAATATCGA 2180
573 rTyrCysSerArgGluSerGlnAsnCysAspSerArgAlaAspAlaGlyP 590
2181 CATTTCCCATCATCTCTGGCGCGTGTGACCAGAAAGTAGTCTATGA 2230
590 heMetValAspValTyrAspIleSerProGlnGlnValAlaGluAlaMet 606
2231 TGATGTGCAGGTGTATGACGCGTCCCTTTATGAGATATGTGGCGCATG 2280
607 AspLysLysGlyAlaLeuValPheAspIleAlaLeuMetPheProValGI 623
2281 ATCAAGAAGAAAGCGCATAACTACTTAACCATGCTACGCCCGCGCA 2330
623 uLeuLeuTyrGlyAsnGlyGluValTyrLeuGluLeuAspThrLeuV 640
2331 GTTTCCTGACGCGCGCAATGCGTCTACATGGAGTCTGTAGACTGTGAGA 2380
640 allLysArgGlu.....GlyAspTyrLeuAlaTyrAsnValGlyGlnCys 654
2381 TTGAAGTGTATGTCGACGCGCGAGTCTGTAATGTACAAATTCGGT..... 2424
655 GlyGluMetTyrGluHisSerPheSerAsnValSerGlyPhePheThrPh 671
2425 AGTTCTTGCTATTCGCACAGCTTTCATCATCAAGGACATCATGACCAC 2474
671 eSerTyrValArgThrSerSerGlyAsnValPheLysLeuGluTyrGluG 688
2475 TCCGTACTTG...ACACTAGTGTGTTTCTATTTCAGCGTGGAGATGTATG 2521
688 lyTyrArgCysGlyTyrHisHisLeuThrMetCysArgAlaGlnLysSer 704
2522 AGTGGGTATGGCGGTGATTTCAAGATTAGGAAGTCCGAAGTATCG 2571
705 ProGlyThrGluValThr.....TyrArgSerLeu..... 714
2572 CCTAGCATTAAGTGCACCAAGCTCTCGAGATACCGAAGAGCTAATAGTGA 2621
715ValProSerPheValGlyLysSerLeuValPheI 726
2622 CGTGGTTAAAGTTAAACTTCCACGTTTCGATAAGAAACGTCGATGTGTC 2671
726 leProValValAlaGlySerSerValSerPheLysThrIleValLeuAsp 742
2672 TGCTT.....GGGTATGACACCATATATACCTAGAT 2700
743 SerAspPheValAspArgIleTyrSerTyrAlaLeuAsnThrIleGlyTh 759
2701 TCGAAGTTTGTGAGTCGCGTTTCGATTATGTCGTGTGTAATTCCTCTGC 2750
759 rPheGluAsnArgThrPheGluTyrAlaValGlyAlaValArgSerGlnL 776
2751 CGTGAACCTCAAAACCTTTCAGTGGGTGTGAGTTTCATTAAAGTCTAGTA 2800
776 yThrHisValIleThrGlySerArgValValHisSerLysValAspIle 792
2801 AGTCGAGGTGATTATTAGCGGTAAATAATTCACAAGGATGTGAATTTG 2850
793 SerProAspAspMetTrpGlyLeuValValAlaValMetAlaGlnAlaI 809
2851 GACCTCAAGTACGTCGAGAGTTTCGCGCGGTTATGTGGCCTCTGCGCT 2900
809 eLysAspArg.....AlaLysSerIleArgSerTyrA 820
2901 GCGCAGTAGACTACGCGTCCGAGTACCTTGTCTAAGAACCTTAGTCATTTT 2950
820 snPheIleLysAlaSerGluGlySerLeuAlaGlyValPheLysLeuPhe 836
2951 CG..... 2952

4543 AACCATGAGTGTTTGGGAAA...CT 4565
 1332 nTyRThrTyrrValGluGluValAlaProSerAlaAlaValAlaGluA 1349
 4566 TTTTGAATTCTA...TCGCTCAAGCTTCACGAAT 4597
 1349 rgGlnGlyAspProSerGlySerGlyThrAlaAspAlaMetAlaPheVal 1365
 4598 TCAGGACTCCGCTTTTGGGAAGATGAACGGAGCTG... 4635
 1366 GluSerValLysLysGlyValAspAsp...ValPheHi 1377
 4636GGAGTTTACGATTTTTCGCCACGGGCATCGGGA 4670
 1377 sGlnGlnSerSerGlyGluThrAlaArgGluValAlaGluValAspGlyLysG 1394
 4671 AAGCGCTCATCGATAGAGAGTGGACCAATTGAA...G 4708
 1394 lyLeuLeuProGluSerValValGlyGluAlaProThrGlnGluArgGly 1410
 4709 AACTTCTCGCGCAGCACCTGAAAGGT... 4734
 1411 ArgAlaAlaAspGlyAsnThrAlaGlnThrAlaValAlaGluGlyAsp... 1426
 4735GACAAGGATGCTTCGTTGACCATGATGAATTACTTTGAGTT 4775
 1427 ArgGluProValGlnSerSerLeuValSerSerProGlnAlaAspIleP 1443
 4776 CTCAGNAGACTTCTTAGGCTCTATCGAGGAGCCGCTTCGCTGGATTCG 4825
 1443 roLysValThrGlnSer... 1448
 4826 GAGGAGGTAGCAAGAACATCGCGATTTTGGCGATTTTGGAAATACGCCAT 4875
 1449GluValHisAlaGlnLysGluValLysGlnGluValProLe 1462
 4876 AATTGTGTTTCGATGTCGCAAGCAAGTGTTCGAACAGCACTTATTCT 4925
 1462 uAlaThrValSerGlyAlaThrProIleValAspGluLysProAlaPro 1479
 4926 TCGTTTCGCCGACCTCTCAGCGCCCTTATCGAGAA...TTTAAGG 4969
 1479 erValThrThrArgGlyValLysIleAspLysGlyLysAlaValAla 1495
 4970 AGGTTTCCCTCGT...AAGACCAAGCTCGTCGCT 5001
 1496 HisValAlaGluLysGlnValGlnValGlnProLysGlnArgse 1512
 5002 ATCGTCGCGAGTATACACTAG...AGATTCTCCGAAGTCGATCGGTGC 5048
 1512 rLeuThrIleAsnGluGlyLysAlaGlyLysGlnLeuLys... 1525
 5049 GTTGGTTTGAATACAGAGTTCGGTGAATAATCTTCGCCGATTTGCTAC 5098
 1526MetPheArgThrCysSerCysGlyValGlnLeuAspValTyr 1539
 5099 CCGCATTAATGAAGCGGAAGTTTCAGTTCGTTCTTTAGCTAGTGTAT 5148
 1540 AsnGlu...AlaThrIleAlaThrArgPhe 1549
 5149 CGCCCACTTAGAGTTTCTCATATATGTGTTCAGCGGAGCGACGTGA 5198
 1549 rAsnAlaPheThrPheValAspAsnLeuLysGlyArgSerAlaValPheP 1566
 5199 AAGTTTGTCTCTCGTG...TGTTTAATCGGGTTAAGTCTCCCTTTCT 5245
 1566 heSerLys... 1568
 5246 TCGTCGCATCGTAGGAGGAACGTCGGAAGAACTCGTGTCTCTCAGCG 5295
 1569LeuGlyGluGlyTyrThr... 1574

5296	CGTCGCTTTATGACGGTATTAAATAATTTTCTTAAGCGCAGAAGTATGCTC	5344
1574	1574
5346	TCATTCTAATTTCTTTTGTCACTTGTTTAGCTCTGACGTTGATGACAGTT	5395
1575TyrAsnGlyGlySerHis.....	1580
5396	CCGCATCTGCAGGGTTGAAAGGTGGTGCATCGCGAATGACGCTCTTCCAC	5445
1581ValSerSerGlyTrpProArgAl	1588
5446	CTTCTGTTGCGCTTGTAGTGCCCTCTATCTAGGTAGGTGG.....	5487
1588	aLeuGluAspIleLeuThrAlaIleLysTyrProSerValPheAspHisC	1605
5488	...GAAAGGTTAAAGCTACTCTTATCGCACCAACTTCATTATTTTGT	5533
1605	ys.....LeuValcInLys.....TyrLysMetGlyGlyGly	1615
5534	GTTTGTGATTGTTGACCATGTGAACGCTCTTATCAAGATTCTTGGGGGT	5583
1616	ValProPheHisAlaAspAspGluCysTyrProSerAspAsnProIl	1632
5584	CTTCTCTTCTTTGTG.....CAACCAAT	5606
1632	eLeuThrVal.....	1635
5607	CTTTTCTGTTTTCGGGGATGCTTCTACAACCGACAGTTTGTGGAGT	5656
1636AsnLeuValGlyLysAlaAsnPheSerThrLysCysArgLys	1649
5657	ATTCGAGAAACTTGTACAGCGTTTGAATTTTCTTAAAGTTCGCGCT	5706
1650	GlyGlyLysValMetValIleAsnValAlaSerGlyAspTyrPheLeuMe	1666
5707	CGCGCGCTGCACCTACTC.....	5724
1666	tProCysGlyPheGlnArgThrHisLeuHisSerValAsnSerIleAspG	1683
5725	AAAGGGTTT.....TTTGAGTCGCTGGCGGA	5749
1683	luGlyArgIleSerLeuThrPheArgAlaThrArgArgValPhe.....	1697
5750	ACACCACTGTGTCAAAAACCGTTCGACACTTCTTCGCTGTTTCGTGAAG	5799
1698GlyValGlyArgMetLeuGlnLeuAlaGly...GlyVa	1709
5800	ATGCTCAAACTTCGAAAAGGCGAGGGTTGCGTGGGATGGTAGGGGTCT	5849
1709	lSerAspGluLysSer...ProGlyValProAsnGlnGlnProGlnSerC	1725
5850	CCATCGCGAAGAACCGTACCCGTCATACCTTCTAATCGGGTGTGACCG	5899
1725	lnGlyAlaThrArgThrIleThrProLysSerGlyGlyLysAlaLeuSer	1741
5900	ACGGGTTGAAAGACTTTCGGTAAAGATGCAAGGACTCAACGGTTCCGT	5949
1742GluGlySe	1744
5950	ACCGAATTGAGAATCTTAGAAGATTTAGATTCTGCGGTGATCGAANAAT	5999
1744	rglyArgGluValLysGlyArgSerThrTyrSerIleTrpCysGluGlnA	1761
6000	CAATAGACGC...AGAAATCGTGACACTAAT.....GACGACG	6034
1761	spTyrValArgLys.....Cys	1766
6035	AATTTACGCGCCTCTCATGAGCAGATGCAAGAAGTCACCACCTTCTGT	6084
1767	GluTrpLeuArgAlaAspAsnProValMetalLeuGlu.....	1779
6085	TCG.....AAAGCCAACCTCTGCTGGTTTGGCCCTTGAAGAGGCAGTGTCT	6128

```
1780 .....ProAspTyrThrPrometThrPheGluV 1789
    : : : : :
6129 TGTGGAGACGCTATAAGTCGGAGAAACTTTCTAAGACGGTTAAATGAGA 6178
    : : : : :
1789 alValIysThrGly.....ThrSerGluAspAlaValVal..... 1800
    : : : : :
6179 TGTGTAGGAAGAGGTACACAGGAGAGAGAGTGGCCGTCGTTGTGTCG 6228
    : : : : :
1801 .....GluTyrLeuLysTyrLeuAlaIleGlyIleGluArg..... 1812
    : : : : :
6229 GACGATGACCGCTGGAAGAAATCTCTGTGCTGACGACGACGACCATTC 6278
    : : : : :
1813 .....ThrTyrArg.....AlaL 1817
    : : : : :
6279 GCCTAAGACAGTCAGGATAAGCGAATACCTTAATAGTTAAACCTCAAGCT 6328
    : : : : :
1817 euLeuMetAlaArgAsnIleAlaValThrThraLaGluGlyValLeuLys 1833
    : : : : :
6329 TCGAATTCGCGAAGCCTATTGTGTGGACGACACAAGGATACCGGGGT 6378
    : : : : :
1834 ValProAsnGlnValTyrGluSerLeuProGlyPheHisValTyrLysSe 1850
    : : : : :
6379 CTAACCAACCGCTGAGGGAG.....TTTTATTATATGCAAGA 6416
    : : : : :
1850 rGlyThrAspLeuIlePheHisSerThr.....GlnAspGlyL 1863
    : : : : :
6417 ACTTGCTCTTTTCGAAATCCACAGCAAACTGTGCACCTACTACGATCAAC 6466
    : : : : :
1863 euArgVal.....ArgAspLeu 1868
    : : : : :
6467 TGGCATAGTCAACTTCGATCGTTCGTAGCACCATGCACGCGAAGTGT 6516
    : : : : :
1869 ProTyrValLeuIleAlaGluLysGlyIlePheThrLysGlyLysAspVa 1885
    : : : : :
6517 CAGCTGTACGTACGGAAGACGCTCAACGATAGTCAGGCTAAACAGGT 6566
    : : : : :
1885 lAspAla....ValValAlaLeuGlyAspAsnLeuPheValCysAspAspI 1901
    : : : : :
6567 ACGTTTGCACATTAAAGGATTCACCGATCAGCATTCCTGTTTGACGGAA 6616
    : : : : :
1901 leu.....ValPheHisAsp 1906
    : : : : :
6617 AAATTTCTATTAAACGCGCGGAGCGGAAATGTTTATATCATCAGAC 6666
    : : : : :
1907 AlaIleAsnLeuIleGlyAlaLeuLysValAlaArgCysGlyMetValG1 1923
    : : : : :
6667 .....AACCTC.....CGGTCTTGGCGAGTAATTTGTTCTAGCCGG 6704
    : : : : :
1923 yGluSerPhe.....LysSerPheGluTyrL 1932
    : : : : :
6705 CTACCCCTTTTCAAGGAGCTTCGTCTTCACGAATTCGTCGGTCGATATTC 6754
    : : : : :
1932 yscystyrAsnAlaProProGlyGlyGlyLysThrThrThrLeuValAsp 1948
    : : : : :
6755 TCCCTTACCAAGATCCACCGGAGGTGGTAAGACGACGACGCTGATGAC 6804
    : : : : :
1949 GluPheVal.....LysSerProAsnSerThrAlaThrIleTh 1961
    : : : : :
6805 TCCTTCTGAAGGCTCTCAAGAAAGGTGAGCTTCCACCATCATCTTAAC 6854
    : : : : :
1961 rAlaAsnValGlySerSerGluAspIleAsnMetAlaValLysLysArga 1978
    : : : : :
6855 CGCCACAAAAGTTCGACGTTGAGATCCTAAAGAAAGTGGAGGAAG 6904
    : : : : :
1978 spProAsnLeuGlu.....GlyLeuAsn 1985
    : : : : :
6905 TGTCTAACATTTGAATGCCAGAACCTAAAGACAAAGATCTCCGAAAAAG 6954
    : : : : :
1986 SerAlaThrThrValAsnSerArgValValAsnPheIleValArgGlyMe 2002
    : : : : :
6955 ACCATTACACCATCGACCTTATTATTATGCAATCAC.....CGTGGTTG 5998
```

```
2002 tTyrLysArgValLeu....ValAspGluValHisMetMetHisGlnGlyL 2018
    : : : : :
6999 TGATGCAGACGCTTTTTCATCGATGAGTGTTCATGGTTCATCGGGTA 7048
    : : : : :
2018 euLeuGlnLeuGlyValPheAlaThrGlyAlaSerGluGlyLeuPhePhe 2034
    : : : : :
7049 GCCTACTAGCTTCGATTCAGTTCACGAGGTGTCATAAAGTAATGATCTTC 7098
    : : : : :
2035 GlyAspIleAsnGlnIleProPheIleAsnArg.....GluLysVa 2048
    : : : : :
7099 GGGGATAGCGGAGATTCTACTACATTGAAGGAACGAATGGACAAAGTG 7148
    : : : : :
2048 lPheArgMetAspCysAlaValPheValProLysLysGluSerValValt 2065
    : : : : :
7149 TTTGTATGGGATCTCGACAGGTTCGTG...GACCTGCAGTGTGGGTTT 7195
    : : : : :
2065 yThrSerLysSerTyrArgCysProLeuAspValCysTyrLeuLeuSer 2081
    : : : : :
7196 ATGGTAATATTTCGATCCGTTGTCATGGGATGTGCGCTTGTGTTAAGC 7245
    : : : : :
2082 SerMet.....ThrValArgGlyThrGluLysCysTy 2092
    : : : : :
7246 ACAGTGTATGGCAACCTAATCGCCACCGTGAAGGTGAAGCGAA..... 7290
    : : : : :
2092 rProGluLysValValSerGlyLysAspLysProValValArgSerLeuS 2109
    : : : : :
7291 .....GGTAAGAGC.....AGCATGC 7306
    : : : : :
2109 erLysArgProIleGlyThrThrAspAspValAlaGluIleAsnAlaAsp 2125
    : : : : :
7307 GCATTAAACGAAATTAATTTCAGTCGACGATTTAGTCCCGACGTGGTTC 7356
    : : : : :
2126 ValTyrLeuCysMetThrGlnLeuGluLysSerAspMetLysArgSerLe 2142
    : : : : :
7357 ACCTTCTGTGTATGCTTCAGTCGGAGAGTTCGAAATCAGCAAGCATT 7406
    : : : : :
2142 uLysGlyLysGly...LysGluThrProValMetThrValHisGluAlaG 2158
    : : : : :
7407 TATTGCAAGGGTTTGACTAAACTTAACGTTCTTAACGGTGCATGAGCGC 7456
    : : : : :
2158 lGlyLysThrPheSerAspValValLeuPheArgThrLysLysAlaAsp 2174
    : : : : :
7457 AAGTGAGACGATATGCGGTGTGAACCTTGTGCGACTTAAGTTTCAGGAG 7506
    : : : : :
2175 AspSerLeuPheThrLysGlnProHisIleLeuValGlyLeuSerArgHi 2191
    : : : : :
7507 GATGAACCTTTAAATCTATCAGGCACATAACCGTCGCTCTTCTCGTCA 7556
    : : : : :
2191 sThrArgSerLeuValTyrAlaAlaLeuSerSerLysLeuAspLysV 2208
    : : : : :
7557 CACCGACAGCTTAACCTATAACGCTTAGCTGCTCGAGGTGACGCCA 7606
    : : : : :
2208 aiclyThrTyrIleSerAspAlaSer 2216
    : : : : :
7607 CTGCGCATGCCATCCAGAGGCTGCG 7632
```

seq_name: /cgnl_8/gcgdata/geneseq/geneseq/NA2000.DAT:A08690

seq_documentation_block:

ID A08690 standard; DNA; 931 BP.

XX A08690;

DT 19-JUL-2000 (first entry)

DE PMwAV-2 ORF 1a DNA encoding a helicase.

KW Pineapple mealybug wilt virus 2; PMwAV-2; ORF 1a; helicase;
KW transgenic pineapple; resistance; antiviral; ss.

OS Pineapple mealybug wilt virus 2.

XX Key Location/Qualifiers

FH


```
FT mat_peptide 2..928
XX /*tag= a
PN WO200017372-A2.
XX
XX 30-MAR-2000.
XX
XX 22-SEP-1999; 99WO-US22152.
XX
XX 23-SEP-1998; 98US-0101461.
XX
XX (UYHA-) UNIV HAWAII.
XX (UYFL ) UNIV FLORIDA.
XX
XX Hu JS, Karasev AV, Dawson WO, Melzer M;
XX
XX WPI; 2000-283596/24.
XX P-PSDB; Y91969.
XX
XX Isolated pineapple mealybug wilt virus proteins and polypeptides,
XX useful for protecting pineapples against the virus
XX
XX Claim 57; Page 96-97; 112pp; English.
XX
XX Pineapple mealybug wilt virus 2 (PMWav-2), open reading frame (ORF) 1a
XX is incomplete but encodes a protein which encompasses all motifs
XX characteristic of viral helicases. The DNA and protein sequences
XX are useful for production of transgenic pineapple plant cultivars and
XX also in transformation methods to impart resistance against the virus to
XX pineapple plants. Mealybug wilt is a major problem limiting profitable
XX pineapple production in many pineapple growing areas worldwide, the
XX present invention aims to overcome this problem.
XX
XX SQ Sequence 931 BP; 262 A; 173 C; 242 G; 254 T; 0 other;

alignment_scores:
  Quality: 631.50 Length: 304
  Ratio: 2.782 Gaps: 4
  Percent Similarity: 74.671 Percent Identity: 44.737

alignment_block:
US-09-301-906-15 x A08690 ..
Align seg 1/1 to: A08690 from: 1 to: 931

1949 GluPheValLysSerProAsnSerThrAlaThrIleThrAlaAsnValGI 1965
||||| ||||| :||||| :||||| :||||| :||||| :|||||
17 GAGTACTTTAAGGATAGATCTGGCAGTGTATTATGACCGCCCAATCGTGG 66

1965 ySerSerGluAspIleAsnMetAlaValLysLysArgAspPro..... 1979
||||| ||||||| :||||| :||||| :||||| :|||||
67 TAGCGGTATAGATATCAATGATACCATTCAGAGTATAGACGCTGCTAATG 116

1980 .....AsnLeuGluGlyLeuAsnSer..... 1986
||||| :||||| :||||| :||||| :|||||
117 CGAGCAAGCCGCTTGAACACAGCTGAGCGGCTGGAATCGATAGATAAT 166

1987 .....AlaThrThrValAsnSerArgValValAlaAsnPhelIleValAr 2000
||||| ||||||| :||||| :||||| :||||| :|||||
167 TATGTTTGTGCTCGTACACTTAACACAAATATGAAAC.....TCCAA 210

2000 gGlyMet.....TyrLysArgValLeuValAspGluValHisMetMeth 2015
||||| ||||| :||||| :||||| :||||| :|||||
211 AGGAGTAATGAATTAACCTGCGCCCTAGTTGACGAAATGTATTGATGC 260

2015 isGlnGlyLeuLeuGlnLeuGlyValPheAlaThrGlyAlaSerGluGly 2031
||||| :||||| :||||| :||||| :||||| :|||||
261 ACAAGGGTCTGTGATGTTGGGGGTATTTCAGTCTGGAGCAAGAGAGCT 310

2032 LeuPheGlyAspIleAsnGlnIleProPheIleAsnArgGluLysVa 2048
||||| :||||| :||||| :||||| :||||| :|||||
311 ATATTCTACGGAGATATCAACCCAGATCCCTTCATATAAATAGAGAAATG 360
```

```
2048 lPheArgMetAspCysAlaValPheValProLysLysGluSerValValt 2065
||||| :||||| :||||| :||||| :||||| :|||||
361 TTTTACTCTAAGGAGGTGTGTACTCTCCAGGTAAGATGAATATTTT 410

2065 yThrSerLysSerTyrArgCysProLeuAspValCysTyrLeuLeuSer 2081
||||| :||||| :||||| :||||| :||||| :|||||
411 ACATATCAGAGTCTTACAGATGCTCCGCCGATGTTGTATGTGGTCAAG 460

2082 SerMetThrValArgGlyThrGluLysCysTyrProGluLysValValse 2098
||||| :||||| :||||| :||||| :||||| :|||||
461 TCATCAAGCGCGCAAGCTGGTCTTAATCGCTACCTGAAGGTGTGTCTG 510

2098 rGlyLysAspLysProValValArgSerLeuSerLysArgProIleGlyT 2115
||||| :||||| :||||| :||||| :||||| :|||||
511 CAACACGCGTGAAGTGTGTACGAGATTATCCAAACGCCGCTAGTGT 560

2115 hrThrAspValAlaGluIleAsnAlaAspValTyrLeuCysMetThr 2131
||||| :||||| :||||| :||||| :||||| :|||||
561 ACCCGAACAAGTATGATACAAATTCGAAGCTGACGCTTATATAACATTCAAG 610

2132 GlnLeuGluLysSerAspMetLysArgSerLeuLysGlyLysGlyLysG 2148
||||| :||||| :||||| :||||| :||||| :|||||
611 CAGGAGTGTAAAGAAAAGCTCTGAGGGCGCTACGAGCTGTAGCAGGAG 660

2148 uThrProValMetThrValHisGluAlaGlnGlyLysThrPheSerAsp 2165
||||| :||||| :||||| :||||| :||||| :|||||
661 GGATAAGTGTTTACAAAGCATGAGCGCAAGGTATGACTTTTGGCGGG 710

2165 aValLeuPheArgThrLysLysAlaAspSerLeuPheThrLysGln 2181
||||| :||||| :||||| :||||| :||||| :|||||
711 TCGTGTATTGTAGATTAAAGTGCCACGATTCCTGTTTCTCTGAG 760

2182 ProHisIleLeuValGlyLeuSerArgHisThrArgSerLeuValTyrAl 2198
||||| :||||| :||||| :||||| :||||| :|||||
761 CCTCACATTTTAGTTCGACTCTCCACACATACACAATCTGTCTATGC 810

2198 aAlaLeuSerSerLysLeuAspLysValGlyThrTyrIleSerAspA 2215
||||| :||||| :||||| :||||| :||||| :|||||
811 CACTCTAGTAGTAGTTAGCCGCAAGGTAGGTGGCGGCATAGACTCAG 860

2215 laSerProGlnSerValSerAspAlaLeuLeuHisThrPheAlaProAla 2231
||||| :||||| :||||| :||||| :||||| :|||||
861 TTACGCGTAAGGAGGTAAAGTATACGGTACTTAAGACCTTTGTGGCGTCG 910

2232 GlyCysPheArg 2235
||| |||||
911 GCGTTATTTTCCA 922

seq_name: /cgnl_8/gcgdata/geneseq/geneseq/NA2000.DAT:A08698
seq_documentation_block:
ID A08698 standard; DNA; 10000 BP.
XX
XX AC A08698;
XX
XX DT 19-JUL-2000 (first entry)
XX
XX DE Pineapple mealybug wilt virus genome 2.
XX
XX KW Pineapple mealybug wilt virus 2; PMWav-2; transgenic pineapple;
XX resistance; antiviral; ss.
XX
XX OS Pineapple mealybug wilt virus 2.
XX
XX FH Key Location/Qualifiers
XX CDS 1..931
XX FT /*tag= a
XX FT /note= "ORF 1a"
XX FT 879..2558
XX FT /*tag= b
XX FT /note= "ORF 1b"
XX FT 3173..3326
```


DT 07-MAR-2000 (first entry)

DE GLRAV-3 polyprotein proteinase domain DNA.

XX Grapevine; leafroll; grapevine leafroll-associated virus 3; GLRAV-3;

XX viral disease; yield loss; sugar content; inhibition; infection;

KW replication; polyprotein; domain; proteinase; methyltransferase;

KW helicase; RNA-dependent; RNA polymerase; untranslated region;

KW transgenic plant; component; resistant; truncation; deletion; antisense;

KW expression; detection; antibody; ds.

XX Grapevine leafroll-associated virus 3.

OS

XX

FH Key Location/Qualifiers

FT mat_peptide 1..360

FT /tag= a

FT /product= "GLRAV-3 polyprotein proteinase domain"

XX

PN W09555880-A1.

XX

XX

PD 04-NOV-1999.

XX

PF 29-APR-1999; 99WO-US09307.

XX

PR 29-APR-1998; 98US-0083404.

XX

XX (CORR) CORNELL RES FOUND INC.

PA

XX Gonsalves D, Ling K;

PI

XX WPI; 2000-062035/05.

DR

DR P-PSDB; Y58143.

XX

PT Newly isolated grapevine leafroll virus protein or polypeptide useful

PT for producing transgenic plants conferring viral disease resistance -

XX

PS Claim 12; Fig 5; 84pp; English.

XX

CC This sequence represents DNA encoding the proteinase domain of the

CC polyprotein (Y58148) from the grapevine leafroll-associated virus 3

CC (GLRAV-3). Leafroll is a serious viral disease, occurring wherever

CC grapes are grown. Although not lethal, it causes yield losses and

CC reduction in sugar content. The virus encodes several proteins,

CC which may serve as targets for the inhibition of viral infection or

CC replication. These proteins include the 242-248 kb polyprotein (Y58148),

CC encoded by open reading frame (ORF) 1a, and which comprises a proteinase

CC domain (Y58143), a methyltransferase domain (Y58144) and a helicase

CC domain (Y58145); a RNA-dependent RNA polymerase (Y58146), encoded by ORF

CC 1b; and a protein of unspecified function (Y58147), encoded by ORF 1l.

CC Nucleotides encoding these proteins, or fragments thereof, and the 5'

CC and 3' untranslated regions (UTRs) of the genome (Z49201-Z49202) are

CC useful for the generation of transgenic plants and plant components.

CC Such transgenic plants may be resistant to viral disease, for example,

CC this property being conferred on the plants via the use of nucleotides

CC encoding truncated or internally deleted proteins, or via the use of

CC antisense nucleotides to inhibit viral gene expression. The nucleotides

CC may additionally be used in the detection of viral nucleic acids in a

CC tissue sample. The proteins can be used to generate antibodies which

CC may be used to detect GLRAV-3 in plant samples. The isolation of

CC GLRAV-3 facilitates the production of agents that reduce the risk of

CC infection or damage by the virus in vineyards.

XX

SQ Sequence 360 BP; 90 A; 97 C; 100 G; 73 T; 0 other;

alignment_scores:

Quality: 605.00 Length: 120

Ratio: 5.042 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-301-906-15 x Z49203

Align seg 1/1 to: Z49203 from: 1 to: 360

85 ValSerGlySerValSerAlaLeuArgGlyAspGlyLysLysValLeuMe 101

1 GTGAGCGGCTCAGTCAGCGCGCTGAGAGGGGATGGTAAAGAGGTCTTGAT 50

101 tGluAlaArgThrSerThrSerAlaThrSerAspValSerAspPheAspv 118

51 GGAGGCAAGGACCTCACTTCCCGCACTTCCGAGCGTCTCGATTTCGACG 100

118 alValPheGluAlaValSerAsnAlaLeuValValHisTyrHisArg 134

101 TCGTATTTCGAAGCTGTTCTTAATGCATTACTTGTCTGACACTACCCACCG 150

135 ValValProTyrAlaProValLysArgGluGlnProLysProAlaValLy 151

151 GTAGTGCCTGATGCCCGGTCAAGCGCGAGCAGCCTAAACCGGCTGTAA 200

151 sGlnAspGluGlnLysProLysArgGlnAlaSerHisTrpAlaValLysp 168

201 GCAAGATGAGCAGAAGCCCAACGCGCAAGCGTCACATTGGGCTGTTAAGC 250

168 roThrAlaValGlyValHisValProLeuProLysLysGlnGluAlaLeu 184

251 CAACAGCTGTGGCTCCAGTCCACTTCTCTAAAAACAGGAAGCAGCTG 300

185 GluProAlaGlnSerValProGlnGlnSerLeuGluGluLysAlaAlaLe 201

301 GAGCCAGCGCAATCAGTCCCAACACAGCTGTTGGAGGAGGAGCGCGCTT 350

201 uThrPheGly 204

351 GACGTTTGGC 360

seq_name: /cgnl_8/gcgdata/geneseq/geneseq/NA1990.DAT:Q04476

seq_documentation_block:

ID Q04476 standard; CDNA; 3649 BP.

XX Q04476;

DT 05-OCT-1990 (first entry)

DE Sequence derived by reverse transcription of RNA1 of AMV.

XX Alfalfa mosaic virus; AMV; cryptic infection; probe; ss.

XX Nicotiana tabacum L "Xanthi-nc".

FH Key Location/Qualifiers

FT CDS 100..3486

FT /tag= a

FT /note="protein encoded by RNA1"

XX US4921802-A.

XX

PD 01-MAY-1990.

XX

PF 13-JAN-1988; 88US-0144692.

XX

PR 05-MAY-1983; 83US-0492582.

XX

PR 13-JAN-1988; 88US-0144692.

XX

PA (PIOE) PIONEER HI-BRED INTERNATIONAL LTD.

XX

PI Hall TC, Loesch-Fries SL, Jarvis NP, Barker RF;

XX

DR WPI; 1990-171297/22.

XX

DR P-PSDB; R05107.

XX

PT Recombinant cDNA plasmids derived from alfalfa mosaic virus RNA -

PT for detection of cryptic infection by alfalfa mosaic virus.

XX

PS Disclosure: ; P; English.

XX The sequence is that of the bottom (B)- or "RNAI" genome fragment
CC of AMV which is necessary to initiate infection. The sequence can
CC be used as a probe to detect cryptic infection of plants by AMV.
CC See also Q04477 and 78.

XX Sequence 3649 BP; 1039 A; 744 C; 808 G; 1058 T; 0 other;

alignment_scores:

Quality: 330.50 Length: 1871
Ratio: 0.499 Gaps: 66
Percent Similarity: 35.382 Percent Identity: 17.263

alignment_block:

US-09-301-906-15 x Q04476 ..

Align seg 1/1 to: Q04476 from: 1 to: 3649

```
435 ThrLysLeuCysAspMetPheSerGln.....ArgAs 445
||| ||| :||| ||| ||| :|||
217 ACTGCAATCGGAAGAAGTTTTCGAGCGGTCGTCCTATGCCAGGA 266
445 pAlaMetIleArgGluLysProSerHisArgCysAspValPhe..... 459
||||| :||| ||| ||| :||| |||
267 TGCTCTC.....CCTTCAGACAAAGGTGAAGTCTTGAAGATAT 304
460 .....LeuLysProArgGluArgGluLysLeuArgGluLeuPhePro 473
||| ||| :||| ||| ||| :||| |||
305 CCTTTTTCCTGACGCCCGCAACAAACATACACGCGCAACATTCCT 354
474 GluLeuSerIleGlnPheSerAspSerValArgSerHisProPheAl 490
||| ||| :||| ||| ||| :||| |||
355 GGTCCGACGTATTGATTTCACACAGTTCGAGTTCATCTCAGCTTTGC 404
490 aAsnAlaMetArgSerCysPheAsnGlyIlePheSerArgArgCysGlyA 507
||| ||| :||| ||| ||| :||| |||
405 GGCTCCCATCGCTACTACTAGAAACGGAATTTGTTTACCGATGTTT 454
507 snVal.....CysPhePheAspIleGlyGlySerPheThrTyrHisVal 521
||| ||| :||| ||| ||| :||| |||
455 ATACCGGTTGATAGTATTATAGACCTTGGAGGAATTCGTTTCCCATAG 504
522 LysAlaGlyHisValAsnCysHisValCysAsnProValLeuAspVally 538
||| ||| :||| ||| ||| :||| |||
505 AAGGTAAGCGGCATAATGTACATTGCTGCTCCCATATTGGATGCTAG 554
538 sAspVallyArgArgIleAsnGluIleLeuPheLeuSerThrAlaGlyG 555
||| ||| :||| ||| ||| :||| |||
555 AGACGGAGCTAGGCTACAGGAGAGAAATATTGCTCTAAAG..... 594
555 LysSerValSerTyrCysSerArgGluSerGlnAsnCysAspSerArgAl 586
||| ||| :||| ||| ||| :||| |||
595 .....TCGTACGTCGCAAAACACCCGGAATTTGGGGAAGCA..... 633
570 LysSerValSerTyrCysSerArgGluSerGlnAsnCysAspSerArgAl 586
||| ||| :||| ||| ||| :||| |||
634 .....GATTACTGATGACACAGTTCAGAAATGCTCAAGCGGAGC 674
586 aAspAlaGlyPheMetVal...AspValTyrAspIleSerProGlnGlnV 602
||| ||| :||| ||| ||| :||| |||
675 TGCAATATGCTTTCATCCATCTACTAGGATCTCGACGTCGGGAGAT 724
602 aAlaGluAlaMetAspLysLysGlyAlaLeuValPheAspIleAlaLeu 618
||| ||| :||| ||| ||| :||| |||
725 TGGCATGTAGTTTGACCAAAAGCGGTTATGAAATTCATTTGACCATG 774
619 MetPheProValGlu.....LeuLeuTyrGlyAsnGlyGluValty 632
||| ||| :||| ||| ||| :||| |||
775 ATGGTTGATGCAGATGCAGATATGTTGATTTCATAGCGAGGGGAATTC 824
632 rLeuGluLeuAspThrLeuValLysArgGluGlyAspTyrLeuAlaT 649
```

```
825 TAACCTTAATGTTAGATGGAGATCATCATAGAACGATCTCATTCATT 874
649 yTAsn...ValGlyGlnCysGlyGluMetTyrGluHisSerPheSerAsn 664
: ||| ||| :||| ||| ||| :||| |||
875 TCGACTTCATCAGCAGCCCAATTCGGGATATAGTCAATCGGTTTCATTG 924
: ||| ||| :||| ||| ||| :||| |||
665 ValSerGlyPhePheThrPheSerTyrValArgThrSerSerGlyAsnVa 681
: ||| ||| :||| ||| ||| :||| |||
925 TTGAAACACACTATTGACITAC.....AATGC 950
681 lPheLysLeuGluTyrGluGlyTyrArgCysGlyTyrHisHisLeuThrM 698
: ||| ||| :||| ||| ||| :||| |||
951 CGTTGATTGGTGCATGCTGCTTATCGA.....A 979
698 .etCysArgAlaGlnLysSerProGlyThr.....GluValThrTyr 711
: ||| ||| :||| ||| ||| :||| |||
980 TCGAAACGTAAGCAAGATTTTGGTGTGATGGTTATTGACTTAACCTAT 1029
712 ArgSerLeu.....ValProSerPheValGlyLysSerLe 723
||| ||| :||| ||| ||| :||| |||
1030 ...TCCCTTGGATTGTCCCAAGATGCCACACTCCAATGGGAGGTCCTG 1076
723 uValPheIleProValAlaGlySerSerVal..... 734
: ||| ||| :||| ||| ||| :||| |||
1077 CGCCTGGTATAATAGATCAAAAGGACAAATGCTAGTCACACCGTTAACG 1126
735 .....SerPheLysThrIleValLeuAspSerAspPhe 745
||| ||| :||| ||| ||| :||| |||
1127 AGGGTACTATCATCATCATACCAGACAGCAGTAGGCGGGAAGTACTT 1176
746 ValAspArgIleTyrSerTyrAlaLeuAsnThrIleGlyThrPheGluAs 762
||| ||| :||| ||| ||| :||| |||
1177 GTCGATAAGAAAGTCTTACCAGAGTACTGAAGTTGCTTTTAGGCAATT 1226
762 nArgThrPheGluTyrAlaValAlaGlyAlaValArgSerGlnLysThr 777
||| ||| :||| ||| ||| :||| |||
1227 CAGACCTAACGCTGATGCTCATTCGCAATTCAGTCCATAGCGACTATG 1276
778 .....HisValIleThrGlySerArgValValHisSer 788
||| ||| :||| ||| ||| :||| |||
1277 TATCTTCTTCAACGAATCATACCATATCGGTGGTGTGACTCTGATTCG 1326
789 LysValAspIleSerProAspAspMetTrpGlyLeuValValAlaValMe 805
: ||| ||| :||| ||| ||| :||| |||
1327 GSCAAACCTCTCAGCCCGATGACTATATTCAGTAGCAACAACGATTTA 1376
805 tAlaGlnAlaIleLys..... 810
: ||| ||| :||| ||| ||| :||| |||
1377 TTATAGAGTGAAAAAACTCTATAACGCCATTCAGAGATGTTATCCCTCC 1426
811 ..AspArgAlaLysSerIleArgSerTyrAsnPheIleLysAlaSerGlu 826
||| ||| :||| ||| ||| :||| |||
1427 TAGACAAAGGAGAGAGATTCGACTGATGCTGTTTTAAAGGGGCTGAA 1476
827 GlySerLeuAlaGlyValPheLysLeuPhePheGlnThrValGlyAspCy 843
||| ||| :||| ||| ||| :||| |||
1477 GGTCCAACG.....TGGTATTCTGGTCCCTAC 1502
843 sPheSerAsnAlaValSerValTyrAlaLysAlaMetValHisAspAsnP 860
||| ||| :||| ||| ||| :||| |||
1503 CTTTTTAAGTCGCTCGAT.....AAGGTCAATGTTCTCGTGTGATT 1543
860 heAsnValLeuGluThrLeuMetSerMetProArgAlaPheIleArgLys 876
||| ||| :||| ||| ||| :||| |||
1544 TT...GTGCCCAAGCTCTGTGCTGTCCTTAAG..... 1575
877 ValProGlySerValValThrIleCysThrSerGlyAlaSerAspAr 893
1575 ..... 1575
893 gLeuGluLeuArgGlyAlaPheAspIleSerLysGluThrPheGlyArgL 910
||| ||| :||| ||| ||| :||| |||
```

1576AGAG 1579

910 ysLeuLysAsnSerArgLeuArgValPheSerArgAlaIleValGluasp 926
|||||::: ::||| |||||:::

1580 ATTGGAATCTCTATTTC TAGGTCA..... 1605

927 SerIleLysValMeLysAlametyLstThrGluaspGlyLysProLeuPr 943
|||||::: ::||| |||||:::

1606CGCACTTCTCATTTCTGAACGGACACC 1631

943 oileThrGluaspSerValTyraLaPheIleMetGlyAsnValSerAsnv 960
|||||::: ::||| |||||:::

1632 GGTTCAGCACGAGAGCCCGC 1651

960 ahisCystThrArgAlaGlyLeuLeuGlyGlySerLysAlaThrValVal 976
|||||::: ::||| |||||:::

1652 TTCGATGC..... 1659

977 SerSerValSerLysGlyLeuValAlaAargGlyAlaAlaThrLysAlaph 993

1659 1659

993 eSerGlyIleThrSerPheSerThrGlySerLeuPheTyrrAspArg 1010
|||||::: ::||| |||||:::

1660ACAGCGGTCTCTTTAC..... 1677

1010 lylLeuThrGluaspGluArgLeuAspAlaLeuValArgThrGluAsnAla 1026

1677 1677

1027 IleAsnSerProValGlyIleLeuGluThrSerArgValAlaValSerLy 1043
|||||::: ::||| |||||:::

1678CCTATAAGGATGTG.....TTGAA 1697

1043 sValValAlaGlyThrLysGluPheTrpSerGluValSerLeuAsnAspp 1060
|||||::: ::||| |||||:::

1698 ATGCCTAGGAAGTGACAAA.....TTTGAGTCGGTCACATAACTGAT. 1740

1060 heThrThrPheValLeuArgAsnLysValLeuIleGlyIlePheValAla 1076

1740 1740

1077 SerLeuGlyAlaAlaProIleAlaTrpLysTyrrArgArgGlyIleAlaL 1093

1740 1740

1093 aAsnAlaAargTyrrAlaGlySerSerTyrrGluThrLeuSerSerLeus 1110

1740 1740

1110 erSerGlnAlaAlaGlyGlyLeuArgGlyLeuThrSerSerThrValSer 1126
|||||::: ::||| |||||:::

1741CCTAGAGTAACCGGAACACTACCGTGGAT 1770

1127 GlyGlySerLeuValValArgArgGlyPheSerSerAlaValThrValTh 1143
::: ::||| |||||:::

1771TTATACCATCTTTTTCAAAGAAAAATT..... 1797

1143 rArgAlaThrValAlaLysArgGlnValProLeuAlaLeuLeuSerPhes 1160

1797 1797

1160 erThrSerTyraIleSerGlyCysSerMetLeuGlyIleTrpAlaHis 1176

1797 1797

1177 AlaLeuProArgHisLeuMetPhePheGlyLeuGlyThrLeuLeuGl 1193
::: ::||| |||||:::

1798GAAACGGTTTTCTCATTTCTTGGAAAGATTGATGC 1835

1193 yAlaArgAlaSerAlaAsnThrTrpLysPheGlyGlyPheSerAsnAsnt 1210

1836 TCCT..... 1839

1210 rpCysAlaValProGluValValTrpArgGlyLysSerValSerLeu 1221
1839 1839

1227 LeuLeuProIleThrLeuGlyValSerLeuIleIleArgGlyLeuLeuAs 1243
1839 1839

1243 nAspThrIleProGlnLeuAlaTyrValProProValGluGlyArgAsnV 1260
1840TCWCCTCTAATTCCTGATCCAG 1861
1260 alTyrAspGluThrLeuArg.....TyrTyrArgAspPheAspTyrAsp 1274
1862 TATACTTCCAATCACTTGAGGATGTGTAATACTGCTGAATGG..... 1902
1275 GluGlyAlaGlyProSerGlyThrGlnHisGluAlaValProGlyAspAs 1291
1902 1902

1291 pAsnAspGlySerThrSerSerValSerSerTyrAspValValThrAsnV 1308
1903 1903

1903 .CATCAGGAATGCCATTGATCGCTCAAAATAC..... 1935
1308 alArgAspValGlyIleSerThrAsnGlyGluValThrGlyGluGluGlu 1324
1935 1935

1325 ThrHisSerProArgSerValGlnTyrThrTyrValGluGluGluValAl 1341
1935 1935

1341 aProSerAlaAlaValAlaGluArgGlnGlyAspProSerGlySerGlyT 1358
1935 1935

1358 hrAlaAspAlaMetAlaPheValGluSerValLysGlyValAlAspAsp 1374
1936GCGGTACCTGTTAGACGAT 1956
1375 valPheHisGlnGlnSerSerGlyCluThrAlaArgGluValGluValAs 1391
1957 ATCAGGAAGCAGAAAGAGAGAGCTTAAGAGCTAAAGCAAGCAAGTTCGA 2006
1391 pGlyLysGlyLeuLeuProGluSerValGlyGluAlaProThrGlnG 1408
2007 AGATGCTCAAAATTAATAGAGCAATTTG..... 2037
1408 luArgGlyArgAlaAlaAspGlyAsnThrAlaGlnThrAlaValAsnGlu 1424
2037 2037

1425 GlyAspArgGluProValGlnSerSerLeuValSerSerProGlnAlaAs 1441
2037 2037

1441 pIleProLysValThrGlnSerGluValHisAlaGlnLysGluValLysG 1458
2038CAAGTTCATGCCTATTGGAAGTCAATC 2065
1458 lnGluValProLeuAlaThrValSerGlyAlaThrProIleValAspGlu 1474
2066 CAGAT..... 2070

1475 LysProAlaProSerValThrArgGlyValLysIleIleAspLysG 1491
2071GCAAGAAAAATCGAAGGACTGGG 2093
1491 yLysAlaValAlaHisValAlaGluLysLysGlnValGlnValGluGlnP 1508
2094 GTTGAGTTCACAGTTCACGAAAAAATCCCGAGCTTCGAATTCACGAC 2143

1508	roLysGlnArgSerLeuThrIleAsnGluGlyLysAlaGlyLysGlnLeu	1524
2144	CAAAACCGCTTACCTGAATTCGAGAGACGACAGAAATGCGCAAAATTTG	2193
1525	CysMetPheArgThrCysSerCysGlyValGlnLeuAspValTyrAsnG1	1541
2194	CGATATCAATCCCTCATTCAGA	2213
1541	uAlaThrIleAlaThrArgPheSerAsnAlaPheThrPheValAspAsnL	1558
2214	TGCCATTCTTGAGCA	2239
1558	euLysGlyArgSerAlaValPhePheSerLysLeuGlyGluGlyTyrThr	1574
2240	TGAAGTCCTACTTCAGCC	2256
1575	TyrAsnGlyGlySerHisValSerSerGlyTyrProArgAlaLeuGluAs	1591
2256		2256
1591	pileLeuThrAlaIleLysTyrProSerValPheAspHisCysLeuValG	1608
2256		2256
1608	lnLysTyrLysMetGlyGlyValProPheHisAlaAspAspGluGlu	1624
2256		2256
1625	CysTyrProSerAspAsnProIleLeuThrValAsnLeuValGlyLysAl	1641
2257	AAVTCATTCATTACCTTTGAACAAATTCGGT	2286
1641	aAsnPheSerThrLysCysArg.LysGlyGlyLysValMetValIleAsn	1657
2287	GATCATTGCTAGTGGACCAAAAGTC	2314
1658	ValAlaSerGlyAspTyrPheLeuMetProCysGlyPheGlnArgThrH	1674
2315	TTGATGTAGTGGG	2329
1674	sLeuHisSerValAsnSerIleAspGluGlyArgIleSerLeuThrPheA	1691
2330	CCGGTCACGACATAAACGTCGAGCTTT	2354
1691	rgAlaThrArgArgValPheGlyValGlyArgMetLeuGlnLeuAlaGly	1707
2355	CATCCCAAGAGAATACTGGGTCGGACCTACTGCTAGAAGTT	2398
1708	GlyValSerAspGluLysSerProGlyValProAsnGlnGlnProGlnSe	1724
2399	ATCCCTTGCAAAATATCAAGA	2421
1724	rgInGlyAlaThrArgThr..IleThrProLysSerGlyGlyLysAlaLe	1740
2422	GCAATGACCAAGGATGATACGTAACTCTGATGGGACGAGAAATGCT	2471
1740	uSerGluGlySerGlyArgGluValLysGlyArgSerThrTyrSerIle.	1756
2472	AGATGCTAATGT	2512
1757	..TrpCysGluGlnAspTyrValArgLysCysGluTrpLeuArgAlaAsp	1772
2513	TCATTGTTGACCA	2550
1773	AsnProValMetAlaLeuGluProAspTyrThrProMetThrPheGluVa	1789
2551	CGPATCATTCCAAGCCCTGGAGAA	2574
1789	lValLysThrGlyThrSerGluAspAlaValValGluTyrLeuLysTyrL	1806
2574		2574
1806	euAlaIleGlyIleGluArgThrTyrArgAlaLeuLeuMetAlaArgAsn	1822

2575	: :	...
	.GCCCTGGGCTTTGAA	..GCACAT 2595
1823	IleAlaValThrThrAlaGluGlyValLeu	...LysValProAs 1836
	: :	:
2596	TTTTCAGTTACGATCTTCATGGAGTCTGCTGGTCCGGAACACCACCAA	2645
1836	n.....GlnValTyrGluSerLeuProGlyPheHisValTyrLysSerG	1851
	: :	:
2646	TATCAAGCAATAGCCGTTCA	...2667
1851	lyThrAspLeuIlePheHisSerThrGlnAspGlyLeuArgValArgAsp	1867
26672667
1868	LeuProTyrValLeuIleAlaGluLysGlyIlePheThrLysGlyLysAs	1884
2668TCGGGTCAAGA	2678
1884	pValAspAlaValAlaLeuGlyAspAsnLeuPheValIcysAspAspi	1901
	: :	:
2679	TGTGGATTGATC.....CTTACCAGCAATCGTAGCTCTGCCGATCAGT	2722
1901	IleuValIlePheHisAspAlaIleAsnLeuIleGlyAlaLeuLysValAla	1917
	: :	:
2723	TAAAGAACAACCATCGATTGTTACCGGTTGACAAAGTTGCATTACATTCGT	2772
1918	ArgCysGlyMetValGlyGluSerPheLysSerPheGluTyrLysCyst	1934
2773	ACCTGT.....	...2778
1934	rAsnAlaProProGlyGlyGlyLysThrThrLeuValAspGluPheV	1951
2779GATCTTAC	2787
1951	aLysSerProAsnSerThrAlaThrIleThrAlaAsnValGlySerSer	1967
27872787
1968	GluAspIleAsnMetAlaValLysLysArgAspProAsnLeuGluGlyLe	1984
2788TT	2789
1984	uAsnSerAlaThrThrValAsnSerArgValValAsnPheIleValArg	2001
2790	GATGCTCCCTCGCGGTAAAGCA	2814
2001	lyMetTyrLysArgValLeuValAspGluValHisMetHisGlnGly	2017
2815CAGAGGTTAATCTTTGATGAATGTTTTTGCACATGCAGCT	2856
2018	LeuGluGlnLeuGlyValPheAlaThrGlyAlaSerGluGlyLeuPheph	2034
2857	TTAGTCTATGCGCTGCTACTTTAGCTGGTGTGTAGCGCAATGTTGGT	2906
2034	eGlyAspIleAsnGlnIleProPheIleAsnArgGluLysValPheArgM	2051
2907	TGGTGACACGGAGCAATCTCTTTGTCTCAGGAATCCGTCATTGGTT	2956
2051	eAspCysAlaValPheValProLysLysGluSerValValTyrThrSer	2067
2957	TTGCTGTCATCAAGCTAAGTGGGAAAGTCGAGAGAAATTAATTACC	3003
2068	LysSerTyrArgCysProLeuAspValCysTyrLeuLeuSerSerMeth	2084
3004TGGAGATCCCGAGAGATGCCACCTATTGCGCTT	3036
2084	rValArgGlyThrGluLysCysTyrProGluLysValValSerGlyLysA	2101
3037GAAAGTATTTTTCAGAACCA	3058
2101	sPlysPro.....ValValArgSerLeuSerLysArg	2111

```
3059 AGAACCGGTGAAGACGAATTCAGAGTACTAAGATCTATCGAAGTTGTG 3108
2112 ProIleClyThrThrAspValAlaGluIleAsnAlaAspValTyrLe 2128
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3109 CCGATAAATCCCTGTAAGCGTTGAGAGAAATACCAACGCTCTTTATTT 3158
2128 uCysMetThrGlnLeuGluLysSerAspMetLysArgSerLeuLysGlyL 2145
||||| ||||||| |||||||:|||||:|||||:|||||:|||||:|||||
3159 GTGTCTACTACAGCTGAAGAGCAGTTTGAAGCTCAACACATTTAA 3208
2145 ysglyLysGluThrProValMetThrValHisGluAlaGlnGlyLysThr 2161
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3209 AGGAGTGTGATAAT...ATCTTACTACTCATCAAGCTCAGGCTAAGACT 3255
2162 PheSerAspValValLeuPheArgThrLysLysAlaAspSerLeuPhe 2178
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3256 TTCGACAATGTTTATTCTGTCGTTAACTCGTACCTCAACGAGTCTTAC 3305
2178 eThr...LysGlnPro.....HisIleLeuValGlyLeuS 2189
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3306 TACTGTGATAGATCCATAAATAGCCCTGATGATGATGATGATGATGAT 3355
2189 erArgHisThrArgSerLeuValTyrAlaAlaLeuSerLysLysLeuAsp 2205
||||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3356 CGAGACACAAGAAGACTTTTAAATATTTTACCACATCGCCCATGATGAT 3405
2206 AspLysValGlyThrTyrLysSerAspAlaSerProGlnSerValSerAs 2222
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3406 GATGTGATCTACATGCTTTAGAGATGCGGT.....AATACCGACGA 3449
2222 pAlaLeuLeu 2225
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3450 TAGTATTTTA 3459
```

seq_name: /cgn1_8/gcgdata/geneseq/geneseq/NA2000.DAT:A08697

seq_documentation_block:

ID A08697 standard; DNA; 5217 BP.

AC A08697;

DF 19-JUL-2000 (first entry)

DE Pineapple mealybug wilt virus genome 1.

KW Pineapple mealybug wilt virus 1; PMWaV-1; transgenic pineapple;

KW resistance; antiviral; ss.

OS Pineapple mealybug wilt virus 1.

FH Key Location/Qualifiers

FT prim_transcript 1..1799

FT /*tag= a

FT /*note= "contains ORF 1a"

FT prim_transcript 1783..3350

FT /*tag= b

FT /*note= "contains ORF 1b"

FT prim_transcript 3360..3499

FT /*tag= c

FT /*note= "contains ORF 2"

FT CDS 3183..5012

FT /*tag= d

FT /*note= "ORF 3"

PN WO200017372-A2.

XX 30-MAR-2000.

XX 22-SEP-1999; 99WO-US22152.

XX 23-SEP-1998; 98US-0101461.

XX (UYHA-) UNIV HAWAII.

PA

(UYFL) UNIV FLORIDA.

PI Hu JS, Karasev AV, Dawson WO, Melzer M;

XX WPI; 2000-283596/24.

DR P-PSDB; Y91963, Y91964, Y91965, Y91966.

XX Isolated pineapple mealybug wilt virus proteins and polypeptides,

PT useful for protecting pineapples against the virus

XX Claim 47; Page 73-75; 112pp; English.

XX This is the pineapple mealybug wilt virus 1 (PMWaV-1) genome which spans

CC 4 open reading frames which may encode four protein products. The DNA

CC and protein sequences are useful for production of transgenic pineapple

CC plant cultivars and also in transformation methods to impart resistance

CC against the virus to pineapple plants. Mealybug wilt is a major problem

CC limiting profitable pineapple production in many pineapple growing areas

CC worldwide, the present invention aims to overcome this problem.

XX Sequence 5217 BP; 1565 A; 925 C; 1267 G; 1460 T; 0 other;

alignment_scores:

Quality: 324.00 Length: 521

Ratio: 1.133 Gaps: 23

Percent Similarity: 54.894 Percent Identity: 26.104

alignment_block:

US-09-301-906-15 x A08697

Align seg 1/1 to: A08697 from: 1 to: 5217

1787 PheGluValValLysThr.....GlyThrSerGluAs 1797

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

351 TTTGAAGTTATCAAGACTATCTAGCCAATACTCTTCGATTTCATTAC 400

1797 pAlaValValGluTyrLeuLysTyrLeuAlaIleGlyIleGluArgThrT 1814

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

401 TGAATGACACAGTTTCTAGACATA.....GAAAGTT 432

1814 YrArgAlaLeuLeuMetAlaArgAsnIleAlaValThrThrAlaGluGly 1830

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

433 ATAGGACACTGGTTAGG.....AGAGTAAGTGGATCGATGGG 470

1831 ValLeuLysValProAsnGlnValTyrGluSerLeuProGlyPheHis.. 1846

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

471 TATTCACCCCTATTGGAAGCTTTATGTCATTCGCTGCACGGTTGAGGGC 520

1847ValTyrLysSerGlyThrAspLeuIleP 1856

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

521 TGAGGTTGAGCTTTTGACAAAGTTGTTAAACAGGAGATCCCGTTAATCG 570

1856 heHis.....SerThrGlnAspGlyLeuArgVal 1865

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

571 GTCACAAACACGAATCTTTGTAAGAACGTCACAAAGAGTCGATAGAGTA 620

1866 ArgAspLeuPro..... 1869

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

621 AAGTACAAACCGGACGGTTCCTATTATACACGTAGAGCTCAGCAACAT 670

1870 TyrValLeuIle...AlaGluLysGlyIlePheThrLysGlyLysAspV 1885

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

671 TTACGGTTTAGTGAACTCACAAGTTGGTATACACAAC..... 710

1885 aAlaValValAlaLeuGlyAspAsnLeuPheValCysAspAspIle 1901

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

711 ..GATCTGTGGTTAAAGAAATCGGAGGGATAGTTTGTAGAACCA... 755

1902 LeuValPheHisasp...AlaIleAsnLeuIleGlyAlaLeu..... 1914

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

756CATGAAATCTCGTTCAATTTAATCCGATCGTTAGCGTTAAT 796

```
1915 .....LysValAlaArgCysGlyMetValGlyGluSerPheL 1927
      |||||:
797 AGACCTTCCTGGTGAAGTTAGC.....CTCGAGGAAGTTAATA 834
1927 ysSerPheGlu...TyrLysCysTyrAsnAlaProProGlyGlyLys 1942
      ||:|||||
835 AGCGGATTGACAATGTGAANTTTGCAACGCTGTTCCAGGGCAGGGAAA 884
1943 ThrThrThrLeuValAspGluPheVal.....LysSe 1953
      |||:
885 ACCTACCAGATCAACACAGAGAATGTTACGCTGCTCGATTCCGAAAGGA 934
1953 rProAsnSerThrAlaThrIleThrAlaAsnValGlySerSerGluAspI 1970
      ||:|||||
935 TGGTAGTCGCTGCTTTAGTCTTCATCGAGTAGAAGTTCAGCTGATACGC 984
1970 lcaSn....MetAlaValLysLysArgAspProAsnLeuGluGluLeuAsn 1985
      ||:|||||
985 TAAAAGCTTTCGCACAGGAGGAGG.....TTGGGCAAGATGATT 1025
1986 SerAlaThrThrValAsnSerArgValValAsnPheIleValArgGly.. 2001
      ||:|||||
1026 CAGATTTTAACGCTCGATGCAATTTTG.....TTCCACGAGAGGAAG 1069
2002 .....MetTyrLysArgValLeuValAspGluValHisMetMeth 2015
      ||:|||||
1070 GAATGTAAGACTTTATAAACGCTTCTGATGACCAATGCTACATGACAC 1119
2015 isGlnGlyLeuLeuGlnLeuGlyValPheAlaThrGlyAlaSerGluGly 2031
      || |||||:|||||
1120 ACGCGGGCATATTGCGAGGTATAAATCGCGCTGTGAAGCCGGAAGAGTGT 1169
2032 LeuPhePheGlyAspIleAsnGlnIleProPheIleAsnArgGluLysVa 2048
      ||:|||||
1170 GTACTTTACGTTAGTATGCGGACAAAGTCCCTTTATCAACAGAAATTA 1219
2048 lPheArgMetAspCysAlaValPheValProLysLysLysGluSerValVal 2065
      ||:|||||
1220 GCTAAATGACAAACAAATCTTTTAAACCGAGTCTCGGCAATTTATCAG 1269
2065 yrThrSerLysSerTyrArgCysProLeuAspValCysTyrLeuLeuSer 2081
      ||:|||||
1270 AGATGTTGATAACAGCGGGTGTCCCGACAGCATCTGTGGCGGATGCC 1319
2082 SerMet....ThrValArgGlyThrGluLysCysTyr.....ProGluLys 2095
      ||:|||||
1320 AATGTTAACTAGGGAAGAGAGATCGTTGTACTCAGGACCGGTAAA 1369
2095 sValValSerGlyLysAspLysProValValArgSerLeuSerLysArgP 2112
      ||:|||||
1370 ATTGTTTACA...CAGTCTAAACAGTCTCTAAAGTGTAAACATCTAAGG 1416
2112 roileGlyThrThrAspAspValAlaGluIleAsnAlaAspValTyrLeu 2128
      ||:|||||
1417 CATTCAAGTAAAGGTGATCACAATCTATTCTCACAAGTGGACAGGTCATG 1466
2129 CysMetThrGlnLeuGluLysSerAspMetLysArgSerLeuLysGlyLys 2145
      ||:|||||
1467 ACCTTTACTCAGACGACGACGAAGTAATCACTCTCAGAGTACATCAGGAG 1516
2145 sGly.....LysGluThrProValMetThrValHisG 2156
      ||:|||||
1517 GCGAATAGGACTATACAAGATGCCAAAACCTTGTATCGGTACGGTTGCTG 1566
2156 luAlaGlnGlyLysThrPheSerAspValValLeuPheArgThrLysLys 2172
      ||:|||||
1567 AGTCCCAAGGTGAAGTACTACAGAGAGTTCATTTTAGTGGCATTTCAACCA 1616
2173 AlaAspAspSerLeuPheGlnProHisIleLeuValGlyLeuSe 2189
      ||:|||||
1617 ACAGACGATCAAGTATTCTTCATGCGCACATAGTACAGCCCTATC 1666
2189 rArgHisThrArgSerLeuValTyrAlaAlaLeuSerSerLysLeuAspA 2206
```

```
|||||
1667 GAGACACACTATCAGCTCAATACTTTTGTATACCAATAAGATGAACA 1716
2206 splysValGlyThrTyrIleSerAsp..... 2214
      ||:|||||
1717 AGGGCATAGGTGAGGATGTACATACTCAATAATCAAGCTGGAGGAAAGAGTC 1766
      ||:|||||
2215 AlaSerProGlnSerValSerAspAlaLeuLeuHisThrPheAlaProAl 2231
      ||:|||||
1767 GCTGCGAATTTTGTACTCCAGCAGTCGCTTAAACATTTATCTCCGTGTA 1816
2231 aGlyCysPheArg 2235
      |||||
1817 AGGATGTGGTGGT 1829
seq_name: /cgnl_8/gcgdata/geneseq/geneseq/NA2000.DAT:A08684
seq_documentation_block:
ID A08684 standard; DNA; 1799 BP.
XX
AC A08684;
XX
DT 19-JUL-2000 (first entry)
XX
DE PMWav-1 helicase DNA.
XX
KW Pineapple mealybug wilt virus 1; PMWav-1; ORF 1a; helicase;
KW transgenic pineapple; resistance; antiviral; ss.
XX
OS Pineapple mealybug wilt virus 1.
XX
FH Key Location/Qualifiers
FT mat_peptide 3..1799
FT /tag= a
FT /product= helicase
XX
PN WO200017372-A2.
XX
PD 30-MAR-2000.
XX
PF 22-SEP-1999; 99WO-US22152.
XX
PR 23-SEP-1998; 98US-0101461.
XX
PA (UYHA-) UNIV HAWAII.
PA (UYFL) UNIV FLORIDA.
XX
PI Hu JS, Karasev AV, Dawson WO, Melzer M;
XX
DR WPI; 2000-283596/24.
DR P-PSDB; Y91963.
XX
PT Isolated pineapple mealybug wilt virus proteins and polypeptides,
PT useful for protecting pineapples against the virus
XX
PS Claim 38; Page 79; 112pp; English.
XX
CC Pineapple mealybug wilt virus 1 (PMWav-1), open reading frame (ORF) 1a
CC encodes a protein with all 8 motifs conserved in the so-called viral
CC helicases. The DNA sequences and related proteins are useful for
CC production of transgenic pineapple plant cultivars and also in
CC transformation methods to impart resistance against the virus to
CC pineapple plants. Mealybug wilt is a major problem limiting profitable
CC pineapple production in many pineapple growing areas worldwide, the
CC present invention aims to overcome this problem.
XX
SQ Sequence 1799 BP; 565 A; 323 C; 455 G; 456 T; 0 other;
```

```
alignment_scores:
  Quality: 323.00      Length: 487
  Ratio: 1.192        Gaps: 22
Percent Similarity: 55.647 Percent Identity: 26.694
```


alignment_block:

US-09-301-906-15 x A08684

Align seg 1/1 to: A08684 from: 1 to: 1799

```

1787 PheGluValValLysThr.....GlyThrSerGluAs 1797
|||||:|||||
351 TTTGAAGTTATCAAGACTATCCTAGCCAATACTCTTGGATTTTCATTAC 400
|||||:|||||
1797 pAlaValValGluTyrLeuLysTyrLeuAlaIleGlyIleGluArgThrT 1814
|||||:|||||
401 TCGAATGAACAGTTTCTAGACATA.....GAAAGTT 432
|||||:|||||
1814 yArgAlaLeuLeuMetAlaArgAsnIleAlaValThrThrAlaGluGly 1830
|||||:|||||
433 ATAGGACACTGGTAGG.....AGAGTAAGTGGATCTGATGGG 470
|||||:|||||
1831 ValLeuLysValProAsnGlnValTyrGluSerLeuProGlyPheHis.. 1846
|||||:|||||
471 TATTCAGCGGTATTGGAAGCTTTATGTCATTCGCTGCACGGGTGAGGCG 520
|||||:|||||
1847 .....ValTyrLysSerGlyThrAspLeuLeuP 1856
|||||:|||||
521 TGAGGTTGACGTTTGGACAAAGTTGTTAAACAGAGAGATCCCGTTAATCG 570
|||||:|||||
1856 heHis.....SerThrClnAspGlyLeuArgVal 1865
|||||:|||||
571 GTCACAAACACGANTGTTTGTGAAGACGTACAGAGCTGGATAGATGA 620
|||||:|||||
1866 ArgAspLeuPro..... 1869
|||||
621 AAGTACAAACCGGCGGGTCCATTATTACACGTAGAGCTCAGCAACAT 670
|||||
1870 TyrValLeuIle...AlaGluLysGlyIlePheThrLysGlyLysAspV 1885
|||||:|||||
671 TTACGGTTTGTAGTCAACACTCACAAGTTGTTATACAAAC..... 710
|||||
1885 alAspAlaValValAlaLeuGlyAspAsnLeuPheValCysAspIle 1901
|||||:|||||
711 ..GATCCTGTGGTTAAAGANTCGGAGGGGTAGTTTGTAGAACCA... 755
|||||
1902 LeuValPheHisAsp...AlaIleAsnLeuIleGlyAlaLeu..... 1914
|||||:|||||
756 .....CATGAATCTCGTTCATTTAATCCGATCGTTAGCGTTAAT 796
|||||
1915 .....LysValAlaArgCysGlyMetValGlyGluSerPheL 1927
|||||:|||||
797 AGACCTTCTGTTGTAAGAGTTAGC.....GTCGAGGAAGTTAATA 834
|||||
1927 ySerPheGlu...TyrLysCysTyrAsnAlaProProGlyGlyLys 1942
|||||:|||||
835 AGCGGATTCACANTGTGAATTTGTGAACGCTGTTCAGGGGCGAGGAAA 884
|||||
1943 ThrThrThrLeuValAspGluPheVal.....LysSe 1953
|||||:|||||
885 ACCTACCATCATCAACAGAGAATGTTACCTGTTTCGATTCGGAAAGGA 934
|||||
1953 rProAsnSerThrAlaThrIleThrAlaAsnValGlySerSerGluAspI 1970
|||||:|||||
935 TGGTAGTGGCTTGTAGTCTGACTTCGAGTACAGAACTCAGCTGATACGC 984
|||||
1970 leAsn...MetAlaValLysLysArgAspProAsnLeuGluGlyLeuAsn 1985
|||||:|||||
985 TAAAGCTTTCGCACAGGAGAGAGG.....TTGGCGAAGATGATT 1025
|||||
1986 SerAlaThrThrValAsnSerArgValValAsnPheIleValArgGly.. 2001
|||||:|||||
1026 CAGATTTTAAACGTCGATCATTG.....TTCCAGCGAGAGGAAG 1069
|||||
2002 .....MetTyrLysArgValLeuValAspGluValHisMetMeth 2015
|||||:|||||
1070 GAATGTGAACACTTTATAAAACCTGCTGATAGACGAATGCTACATGACAC 1119

```

```

2015 isGlnGlyLeuLeuGlnLeuGlyValPheAlaThrGlyAlaSerGluGly 2031
|||||:|||||
1120 AC CGCGGCATATTGCGAGGTATTAATCGCCCTGTGAAGCGGAGAGTGT 1169
|||||:|||||
2032 LeuPhePheGlyAspIleAsnGlnIleProPheIleAsnArgGluLysVa 2048
|||||:|||||
1170 GTACTTTACGGTGATAGGGGACAAAGTCCCTTTATCAACAGAATCAAA 1219
|||||:|||||
2048 lPheArgMetAspCysAlaValPheValProLysLysGluSerValVal 2065
|||||:|||||
1220 GCTAAATGACAAACAATCCTTTTGAACCGGAGTCTCGGCAATATTTC 1269
|||||:|||||
2065 yThrSerLysSerTyrArgCysProLeuAspValCysTyrLeuLeuSe 2081
|||||:|||||
1270 AGATGTTGATAACCGGGTGTCCCGCAGACATCTGTTGGCGGATGTC 1319
|||||:|||||
2082 SerMet...ThrValArgGlyThrGluLysCysTyr.....ProGluLy 2095
|||||:|||||
1320 AATGTTTAAACAATGGGAAGAAAGAGATCGTTTGTACTCAGGACCGGT 1369
|||||:|||||
2095 sValValSerGlyLysAspLysProValValArgSerLeuSerLysArgP 2112
|||||:|||||
1370 ATTGTTTACA...CAGTCTAAACCAAGTCTCTAAAAGTGAACATGTA 1416
|||||:|||||
2112 roileGlyThrThrAspAspValAlaGluIleAsnAlaAspValTyrLe 2128
|||||:|||||
1417 CGTTCAGTAAAGTGATCACAATCTATTCTCACAGTGGACAGGGTCAT 1466
|||||:|||||
2129 CysMetThrClnLeuGlySerSerAspMetLysArgSerLeuLysGly 2145
|||||:|||||
1467 ACCTTTACTCAGAACGAGAAGAAATGAATCATCTCAGAGTACATGAG 1516
|||||:|||||
2145 sGly.....LysGluThrProValMetThrValHisG 2156
|||||:|||||
1517 GGAATAGGAGCTATACAAAGATCGAAAACTTTGACGGTACGGTGTG 1566
|||||:|||||
2156 luAlaGlnGlyLysThrPheSerAspValValLeuPheArgThrLysLys 2172
|||||:|||||
1567 AGTCCCAAGGTGAACCTTACAGAGAGTTCATTAGTGGCATTCAACAC 1616
|||||:|||||
2173 AlaAspAspSerLeuPheThrLysGlnProHisIleLeuValGlyLeuSe 2189
|||||:|||||
1617 ACAGACGATCAAGTATTTCTTCCATGCCACATAGACTAGTACCCCTAT 1666
|||||:|||||
2189 rArgHisThrArgSerLeuValTyrAlaAlaLeuSerSerLysLeuAspA 2206
|||||:|||||
1667 GAGACACACTATCAGCTGCAATACTTTTGTATACCCAATAAGATGAAC 1716
|||||:|||||
2206 sPlysValGly 2209
|||||:|||||
1717 AGGGCATAGGT 1727

```

seq_name: /cgnl_8/gcgdata/geneseq/geneseq/NA1994.DAT:071546

seq_documentation_block:

ID Q71546 standard; cDNA; 6384 BP.

XX Q71546;

XX AC Q71546;

XX DT 21-JUN-1995 (first entry)

XX DE Component of transcriptional vector of tobamovirus.

XX KW Tobamovirus; tobacco mosaic virus; vector; transcription; promoter;

XX KW terminator; replication enzyme; replication site; ori; coat protein;

XX OS Tobacco mosaic virus.

XX FH Key Location/Qualifiers

XX FT CDS 72..3422

FT /*tag= a

744 spPheValAspArgIleTyrSerTyr.....AlaLeuAsnThrIleGly 758
:::||||| ::::: ||| ::::: |||
966 ..TTAGTAAGTAACTAGTAACTGTTTGTAAATTTTCTAGAAATAGAT 1013
:::||||| ::::: ||| ::::: |||
759 ThrPheGluAsnArgThrPheGluTyrAlaValGlyAlaValArgSerGI 775
|||||
1014 ACTTTC..... 1019
775 nLysThrHisValIleThrGlySerArgValValHisSerLysValAspI 792
::: ||| ::::: ||| ::::: |||
1020TTATTGTACAAAGGTGTAGCGCATAGGGTGTAGAT. 1055
::: ||| ::::: ||| ::::: |||
792 leSerProAspMetTrpGlyLeuValValAlaValMetAlaGlnAla 808
1055 1055
809 IleLysAspArgAlaLysSerIleArgSerTyrAsnPheIleLysAlaSe 825
||| ::::: ||| ::::: |||
1056AGTGACGAGTTTACNAGGCTAT 1078
825 rGluGlySerLeuAlaGlyValPheLysLeuPheGlnThrValGlyA 842
||| ::::: ||| ::::: |||
1079 GGAAGACGATGCGAC.....TACAAAAAGACTCTTGCGA 1113
842 spCysPheSerAsnAlaValSerValTyrAlaLysAlaMetValHisAsp 858
||| ::::: ||| ::::: |||
1114 TGTGCAACAGTCAAGAANTCTGTTAGAGGATCTTTCATCAGTTAATTAC 1163
||| ::::: ||| ::::: |||
859 AsnPheAsnValIleLeuThrLeuMetSerMetProArgAlaPheIleAr 875
||| ::::: ||| ::::: |||
1164 TGGTTTCCAAAATGAGGATAGGTAGTAGTTCCA..... 1199
875 gLysValProGlySerValValThrIleCysThrSerGlyAlaSerA 892
1199 1199
892 spArgLeuGluLeuArgGlyAlaPheAspIleSerLysGluThrPheGly 908
|||||
1200CTATTTGACATATCTCTCGAGACT..... 1223
909 ArgLysLeuLysAsnSerArgLeuArgValPheSerArgAlaIleValGI 925
||| ::::: ||| ::::: |||
1224AGTAAAGACACGCAAGAGGTCTTA.....GTTTCAAA 1258
925 uAspSerIle...LysValMetLysAlaMetLysThrGluAspGlyLysP 941
||| ::::: ||| ::::: |||
1259 GGACTTTGTTTATACAGTGTAAATCACATTCGTACGTACCAGGCCAAAG 1308
941 roleuProIleThrGluAspSerValTyrAlaPheIleMetGlyAsnVal 957
||| ::::: ||| ::::: |||
1309 CGCTT.....ACTTACTCAACGTTATCTTTCGTCGAATCAATTCTGT 1352
958 SerAsnValHisCysThrArgAlaGlyLeuLeuGlyGlySerLysAlaTh 974
||| ::::: ||| ::::: |||
1353 TCAGAGAGTG.....ATCATTACGGGGTTACTGCTAG 1384
974 rValValSerSerValSerLysGlyLeuValAlaArgGlyAlaAlaThrL 991
::: ||| ::::: ||| ::::: |||
1385 GTCTGAGTGGGATGTCGATAAATCATTTATACAGTCTTGTGCGATGACG. 1433
991 yaAlaPheSerGlyIleThrSerPhePheSerThrGlySerLeuPheTyr 1007
|||||
1434TTCTTC..... 1439
1008 AspArgGlyLeuThrGluAspGluArgLeuAspAlaLeuValArgThrGI 1024
1439 1439
1024 uAsnAlaIleAsnSerProValClyIleLeuGluThrSerArgValAlav 1041
::: ||| ::::: ||| ::::: |||
1440CTACACACCAAGCTTCCCGTTCTGAAAGACGAT...CTTTGA 1479

1041 aISerLysValValAlaGlyThrLys.....GluPheTrpSer 1053
::: ||| ::::: ||| ::::: |||
1480 TTAGCAAGTTTGCACTTGACCAAAACATGCTCACAACATGCTGGGAT 1529
::: ||| ::::: ||| ::::: |||
1054 GluValSerLeuAsnAspPheThrThrPheValLeuArgAsnLysValLe 1070
||| ::::: ||| ::::: |||
1530 GAGATTTCCTTA..... 1541
1070 uIleGlyIlePheValAlaSerLeuGlyAlaAlaProIleAlaTrpLysT 1087
1541 1541
1087 yrArgArgGlyIleAlaAlaAsnAlaArgArgTyrAlaGlySerSerTyr 1103
1541 1541
1104 GluThrLeuSerSerLeuSerSerGlnAlaAlaGlyGlyLeuArgGlyLe 1120
1541 1541
1120 uThrSerSerThrValSerGlySerLeuValValArgArgGlyPheS 1137
::: ||| ::::: ||| ::::: |||
1542GCTTTCG 1548
1137 erSerAlaValThrValThrArgAlaThrValAlaLysArgGlnValPro 1153
::: ||| ::::: ||| ::::: |||
1549 GCAATGCTTTTCCCATCGATCAAGGAAGATTGATAAACCGGAAA..... 1592
1154 LeuAlaLeuLeuSerPheSerThrSerTyrAlaIleSerGlyCysSerMe 1170
1593CT 1594
1170 tLeuGlyIleTrpAlaHisAlaLeuProArgHisLeuMetPhePheG 1187
::: ||| ::::: ||| ::::: |||
1595 GATCAAAATTACGCGAATGCGTTAGAG..... 1622
1187 lyLeuGlyThrLeuLeuGlyAlaArgAlaSerAlaAsnThrTrpLysPhe 1203
1622 1622
1204 GlyGlyPheSerAsnAsnTrpCysAlaValProGlu.....ValValTr 1218
1623ATCAGGGTCCCGCTCTTTATGTCACCTT 1651
1218 pArgGlyLysSerValSerSerLeuLeuProIleThrLeuGlyVals 1235
::: ||| ::::: ||| ::::: |||
1652 CCATGATAGTTAGTTCTGAGTACAAAATGTCAGTGGACATG..... 1694
1235 erLeuIleIleArgGlyLeuLeuAsnAspThrIleProGlnLeuAlaTyr 1251
1694 1694
1252 ValProProValGluGlyArgAsnValTyrAspGluThrLeuArgTyrTy 1268
||| ::::: ||| ::::: |||
1695 ...CCGGTCTAGACATTAGGAAAAAGATGGAAAGAACTAGGAGAAATGTA 1741
1268 rArgAspPheAspTyrAspGluGlyAlaGlyProSerGlyThrGlnHisG 1285
1742 C..... 1742
1285 luAlaValProGlyAspAspAsnAspGlySerThrSerSerValSerSer 1301
::: ||| ::::: ||| ::::: |||
1743AATGCACGTCTCCGAA 1757
1302 TyrAspValValThrAsnValArgAspValGlyIleSerThrAsnGlyGI 1318
::: ||| ::::: ||| ::::: |||
1758 CTGCTGTACTTAAATAATTCAGACAAGTTCGATGTTGACGTTTTTCCCA 1807
1318 uValThrGlyGluGluGluThrHisSerProArgSerValGlnTyrThrT 1335
::: ||| ::::: ||| ::::: |||
1808 GATGTCC.....CAATCTTTA..... 1823
1335 yrValGluGluGluValAlaProSerAlaValAlaGluArgGlnGly 1351

1950 PheValLysSerProAsnSerThrAlaThrIleThrAlaAsnValGlySer 196
2589ACAAGGAAATCTTTCGAGAGTAAATTT 2617
1966 rSerGluAspIleAsnMetAla.....ValLysL 1976
2618 CGAAGAAGACTCTAATTCITGTCCCTGGTCGCAAGCTCCGAGATGATCA 2667
1976 ysArgAspProAsnLeuGluGlyLeuAsnSerAlaThrThrValAsnSer 1992
2668 GAAGAAGAGCTAATCGTCGGGCATACTAGTGGCTACAAAGGATAATGTG 2717
1993 ArgValVal...AsnPheIleVal.....ArgGlyLys 2002
2718 CGACCGTCGATTCAATTTGATGAATTCGGGAAGGGCAGCTGTCA 2767
2002 tThrLysArgValLeuValAspGluValHisMetMetHisGlnGlyLeuL 2019
2768 GTTCAAAAGATTGTTTCATAGCAAGGTTTGATCTCGCATCTGCTGTG 2817
2019 euGlnLeuGlyValPheAlaThrGlyAlaSerGluGlyLeuPhePheGly 2035
2818 TGAATTTTTGGTTGCAATGTCTGTGCGCATATTCGATATGTTATGGA 2867
2036 AspIleAsnGlnIleProPheIleAsnArgGluLysValPheArgMetAs 2052
2868 GACACCCACAGATTCCGTACATCAACAGAGTAATACTGGTTTCCGTACCC 2917
2052 pCysAlaValPheValProLysLysGluSerValValThrSerLys 2069
2918 TGCACACTTTCGAAATTTGGAGGTCGAGGAGTCGAAACGAAGAATA 2967
2069 erTyrArgCysProLeuAspValCysTyrLeuLeuSerSer..... 2082
2968 CTCTTCGTGTGGCGCTGATGTCACACATCTCTTAAATCAAAAGGTAA 3017
2083MetThrValArgGlyThrGluLysCysTyrProGluLysVal 2096
3018 GCACAGCTAATGTGCACGCTCTCTGAAAGAAATTCAGTTTCCACGGAAT 3067
2096 lValSerGly.....LysAspLysProValValArgSerLeuSerLys 2111
3068 GGTTCAGTGGGCTGCGTCTATCAATCTGTCCTCCAGCCGCTTAAGGAA 3117
2111 rgProIleGlyThrThrAspAspValAlaGluIleAsnAlaAspValTyr 2127
3118 AA.....ATT 3122
2128 LeuCysMetThrGlnLeuGluLysSerAspMetLysArgSerLeuLysG 2144
3123 TTGACTTTTACACAG.....TCTGAC...AAGGAGGCCCTTCCTC 3160
2144 yLysGlyLysGluThrProValMetThrValHisGluAlaGlnGlyLys 2161
3161 AAGGGCC...TACGCAGATGTCCTACTGTCATGAGGTACAGGTCAGA 3207
2161 hrPheSerAspValValLeuPheArgThrLysLysAlaAspSerLeu 2177
3208 CTTATGTCAGACGTATCGTTTACGTTTCGCTAACACCTACCGCTGATCATC 3257
2178 PheThrLys...GlnProHisIleLeuValGlyLeuSerArgHisThrAr 2193
3258 ATCCGACAGACAGTCCGCATGTTCTGGTCTGCTGTGTCACAGACACAAA 3307
2193 gSerLeuValTyrAlaAla.....LeuSerSerLysLeuA 2205
3308 ATCCCTAAGTACTACACCGTGTGATGGATCCCTTTAGTTAGTATCATTA 3357
2205 sAsp.....LysValGlyThrTyrIleSerAsp 2214
3358 CAGATTATGAACGGGTAGTAGTTACTTATTATAGAC 3392

```

seq_name: /cgnl_8/gcgdata/geneseq/geneseq/NA1999.DAT:V99284
seq_documentation_block:
ID V99284 standard; cDNA; 8743 BP.
XX V99284;
AC
XX
XX 26-APR-1999 (first entry)
XX
XX Rupestris stem pitting associated virus RSPaV-1 genome.
DE
XX
XX RSPaV-1; grape; transgenic plant; disease resistance; ss.
XX
XX Rupestris stem pitting associated virus.
OS
XX
XX Key Location/Qualifiers
FH 62..6546
FT CDS /*tag= a
FT /*product= replicate
FT /*transl_except= (pos:5852..5853, aa:Asp)
FT /*note= "this codon has an apparent 1 nucleotide
FT deletion which alters the reading frame;
FT this region is specifically claimed in
FT Claim 22"
FT CDS 6577..7242
FT /*tag= b
FT /*product= triple gene block first protein
FT /*note= "this region is specifically claimed in
FT Claim 28"
FT CDS 7244..7597
FT /*tag= c
FT /*product= triple gene block second protein
FT /*note= "this region is specifically claimed in
FT Claim 30"
FT CDS 7519..7760
FT /*tag= d
FT /*product= triple gene block third protein
FT /*note= "this region is specifically claimed in
FT Claim 32"
FT CDS 7770..8549
FT /*tag= e
FT /*product= coat protein
FT /*note= "this region is specifically claimed in
FT Claim 25"
XX WO9852964-A1.
XX 26-NOV-1998.
XX
XX 20-MAY-1998; 98WO-US10391.
XX
XX 17-DEC-1997; 97US-0069902.
XX 20-MAY-1997; 97US-0047147.
XX
XX (CORR ) CORNELL RES FOUND INC.
XX
XX Gonsalves D, Meng B;
PI
XX
XX WPI; 1999-045297/04.
XX P-PSDB; W87724-28.
XX
XX Isolated proteins from Rupestris stem pitting-associated virus a
XX related nucleic acid - vectors, host cells and transgenic Vitis
XX cultivars that are resistant to the virus
XX Claim 63; Page 9-14; 163pp; English.
XX
XX This is the nucleotide sequence of the Rupestris stem pitting
XX associated virus RSPaV-1 genome. It is derived from cDNA clones
XX of a dsRNA that was associated with Rupestris stem pitting. The
XX sequence includes 5 open reading frames (ORF1-5) that encode the
XX viral replicase, the 3 proteins of the RSPaV-1 triple gene block
XX and the coat protein (see W87724-28). Substantial portions of
XX

```

CC genomes of 2 other RSPaV strains, RSP47-4 and RSP158, have also
CC been sequenced (see V99285-86), as have portions of 7 other RSPaV
CC strains (see V99287-93). Also claimed are: an expression system
CC comprising nucleic acids encoding RSPaV proteins or polypeptides;
CC transformed host cells; transgenic grape (Vitis) scion or rootstock
CC cultivars that comprise DNA encoding a replicase, coat protein or
CC triple gene block protein of RSPaV; a method of imparting RSPaV-1
CC resistance to Vitis cultivars by transformation using such DNA;
CC and primers and probes (see V99294-305) for detection of RSPaV
CC sequences.
XX
SQ Sequence 8743 BP; 2430 A; 1677 C; 2076 G; 2560 T; 0 other;

alignment_scores:
Quality: 290.00 Length: 142
Ratio: 2.736 Gaps: 3
Percent Similarity: 74.648 Percent Identity: 41.549

alignment_block:

US-09-301-906-15 x V99284 ..

Align seg 1/1 to: V99284 from: 1 to: 8743

1554 PheValAspAsnLeuLysGlyArgSerAlaValPhePheSerLysLeuCl 1570
||| |||:||||:||||:||||:||||:||||:||||: ||| |||:||||: ||
2258 TTTAAGACAGCGCTCAGAGCGGCTGCATCATTTCTCTCAAAACCAGG 2307
1570 Y...GluGlyTyrThrTyraSnGlyGlySerHisValSerSerGlyTrrp 1586
| :
2308 AATTTCATGTCATAGTTACAAATGGTGGGAGCCACACAAAGCTTAGGGTGGC 2357
1586 roArgAlaLeuGluAspIleLeuThrAlaIleLysTyrProSerValPhe 1602
||:
2358 CAAAGTTCATGCATCAGATTCTTAAGCTCCACTGGTGGAGGTAATTACTAC 2407
1603 AspHisCysLeuValGlnLysTyrLysMetGlyGlyValProPheHi 1619
: : : : |||:||||: |||: : : : : : : : : : : : : : : : : :
2408 AATTTCATGCTGCCTCAGATCTATGAGGAAAATTCAAAATTTGGCTCTTCA 2457
1619 sAlaAspAspGluGluCysTyrProSerAspAsnProIleLeuThrValA 1636
| |||:||||:||||:||||: |||: : : : : : : : : : : : : : : : :
2458 TAAGGATGATGAGAGTGTCTATGAATTTGGGCACAAAGTTTGGACTGTTA 2507
1636 snLeuValGlyLysAlaAsnPheSer...ThrLysCysArgLys..... 1649
|||:||||:||||:||||:||||: |||: : : : : : : : : : : : : : :
2508 ATTTAATCGGCTCAGCAACTTTCACTATTAGTAGTACGCGAAATTTGGTT 2557
1650 GlyGlyLysValMetValIleAsnValAlaSerGlyAspTyrPheLeuMe 1666
|||:||||: ||
2558 GGGGGTAATTCATTGCGCCTGACAATTTGGCCAAAATGAGTTTTTCGAAAT 2607
1666 tProCysGlyPheGlnArgThrHisLeuHisSerValAsnSerIleAspG 1683
||| |||:||||: |||:
2608 GCCTAGGGGCATGCAATGCAATTACTTCCATGGGGTTTCCAAATGTACGC 2657
1683 luGlyArgIleSerLeuThrPheArg 1691
|||:||||:||||:||||:||||: |||: : : : : : : : : : : : : : :
2658 CAGGCGGGGTATCGCTGACCTTTAGG 2683

OM of: US-09-301-906-15 to: GenEmbl.* out_format : pfs
Date: Feb 18, 2001 5:18 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn1/1/USPRO.spool/US09301906/runat_15022001_103218_9315/app_query.fasta_1.2332
-DB=GenEmbl -QFM=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000 -GAPOP=4.500
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -XGAPOP=6.000
-GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000
-DELEX=7.000 -START=1 -MATRIX=blos62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THRCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFM=pfs -NORM=ext -MINLEN=0
-MAXLEN=200000000 -USER=US09301906@cgn1_1.12272 -NCPU=6
-ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-301-906-15
Query length: 2237
Database: GenEmbl.*
Database sequences: 1118133
Database length: 1736092196
Search time (sec): 7258.940000

Sequence	Strid Orig	ZScore	EScore	Len	Documentation
gb_pat1:AR070579	+ 5247.00	4657.66	2.7e-251	4173	! AR070579 Sequence 1 from patent
gb_v11:AF037268	+ 3644.00	3220.08	3.3e-171	13154	! AF037268 Grapevine leafroll-a
gb_v11:AF283103	+ 2233.00	2052.60	3.5e-106	14861	! AF283103 Pineapple mealybug w
gb_v11:AF195822	+ 803.50	693.79	1.7e-30	12394	! AF195822 Grapevine leafroll-a
gb_v11:AF039204	+ 794.50	683.86	6.0e-30	15000	! AF039204 Grapevine leafroll-a
gb_v11:AF190581	+ 767.00	659.09	1.4e-28	15468	! AF190581 Beet yellows virus S
gb_v11:AF055575	+ 762.00	654.64	2.6e-28	15468	! AF055575 Beet yellows virus S
gb_v11:BVUUA	+ 739.00	634.17	3.5e-27	15480	! X73476 Beet yellows virus (is
gb_v11:CTU16304	+ 737.50	630.62	5.6e-27	19296	! U16304 Citrus tristeza virus
gb_v12:CTCVORFS	+ 721.00	617.25	3.1e-26	16934	! Y10237 Little cherry closterc
gb_v11:CTTV18420	+ 711.00	607.06	1.1e-25	19259	! Y18420 Citrus tristeza virus
gb_v11:CTV06051	+ 697.00	594.61	5.7e-25	19259	! AF260651 Citrus tristeza viru
gb_v11:CTAF001623	+ 678.50	578.16	4.7e-24	19249	! AF001623 Citrus tristeza viru
gb_v11:CTU56902	+ 661.50	563.05	3.2e-23	19226	! U56902 Citrus tristeza virus
gb_v12:LTU15440	+ 653.50	564.61	2.7e-23	8118	! U15440 Lettuce infectious yell
gb_v11:AB033689	+ 483.50	414.55	6.1e-15	7226	! AB033689 Soil-borne wheat mos
gb_v11:AB033691	+ 470.00	403.04	2.7e-14	6878	! AB033691 Sorghum chlorotic sp
gb_v11:CWH012005	+ 466.00	399.09	4.4e-14	7147	! AJ012005 Chinese wheat mosaic
gb_v11:DB6636	+ 462.00	397.99	5.1e-14	5600	! DB6636 Broad bean necrosis vir
gb_v12:OG0132578	+ 453.50	388.02	1.8e-13	7111	! AJ132578 Oat golden stripe vir
gb_v11:BSBVRNA1	+ 450.50	387.35	2.0e-13	5834	! Z97873 Beet soil-borne virus
gb_v12:WMORNA1	+ 447.50	382.70	3.6e-13	7099	! L07937 Soil-borne wheat mosaic
gb_v11:AF146278	+ 438.00	374.35	1.0e-12	7026	! AF146278 Soil-borne rye mosaic
gb_v11:AF146280	+ 427.50	365.07	3.5e-12	6985	! AF146280 Soil-borne rye mosaic
gb_v11:EMH132576	+ 423.50	360.46	5.5e-12	7025	! AJ132576 European wheat mosaic
gb_v11:AF146279	+ 422.50	360.62	6.1e-12	6985	! AF146279 Soil-borne rye mosaic
gb_v12:PMO238607	+ 421.00	360.75	6.0e-12	6043	! AJ238607 Potato mop-top virus
gb_v12:PMRNA	+ 412.00	351.77	1.9e-11	6656	! D13957 Papaya mosaic virus gen
gb_v12:SLVRNAGENM	+ 405.00	342.69	6.1e-11	8832	! M97264 Shallot virus X DNA sec
gb_v11:GVBNPCP	+ 393.50	333.98	1.9e-10	7599	! X75448 Grapevine virus B genom
gb_v11:GVMPCP	+ 389.00	330.31	3.0e-10	7349	! X75433 Grapevine virus A genom
gb_v11:BVU51931	+ 386.50	324.45	6.3e-10	10545	! U51931 Beet yellow stunt viru
gb_v11:BVQ223596	+ 373.00	318.11	1.4e-09	6003	! AJ223596 Beet virus Q genomic
gb_v11:AF237816	+ 359.00	302.24	1.1e-08	8432	! AF237816 Cherry necrotic rusty
gb_v11:CYMRNA	+ 349.00	295.20	2.7e-08	7015	! D29630 Clover yellow mosaic po
gb_v12:PNRNA	+ 347.50	294.13	3.1e-08	6828	! L23972 Pepper ringspot viral I
gb_v12:SMU47034	+ 346.50	291.97	4.1e-08	4683	! U47034 Sunn-hemp mosaic virus
gb_v11:AF034622	+ 345.00	291.96	4.1e-08	6790	! AF034622 Tobacco rattle virus
gb_v12:TOTRNA1	+ 345.00	291.96	4.1e-08	6791	! X06155 Tobacco rattle virus RN
gb_v12:AF166084	+ 339.00	286.63	8.1e-08	6791	! AF166084 Tobacco rattle virus

gb_v11:AF017780 + 332.00 278.29 2.4e-07 8372 ! AF017780 Sour cherry green
gb_v11:GLVGNs + 331.00 277.43 2.6e-07 8353 ! Z68502 Garlic latent virus
gb_v12:MAACG12 + 320.00 284.89 1.0e-07 3644 ! L00163 Alfalfa mosaic virus
gb_v11:AB045371 + 320.50 267.01 1.0e-06 9293 ! AB045371 Apple stem pitting
seq_name: gb_pat1:AR070579
seq_documentation_block: 4173 bp DNA PAT 18-FEB-2000
LOCUS AR070579
DEFINITION Sequence 1 from patent US 5907085.
ACCESSION AR070579
VERSION AR070579.1 GI:7221467
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 4173)
AUTHORS Gonsalves,D. and Ling,K.
TITLE Grapevine leafroll virus proteins and their uses
JOURNAL Patent: US 5907085-A 1 25-MAY-1999;
FEATURES
source location/Qualifiers
1..4173 /organism="unknown"
BASE COUNT 1030 a 874 c 1177 g 1092 t
ORIGIN
alignment_scores:
Quality: 5247.00 Length: 1423
Ratio: 4.214 Gaps: 17
Percent Similarity: 87.491 Percent Identity: 74.139
alignment_block:
US-09-301-906-15 x AR070579 ..
Align seg 1/1 to: AR070579 from: 1 to: 4173
848 valSerValTyrAlaLysAlaMetValHisAspAsnPheAsnValLeuL 864
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1 GTGCTACTTACCGAAGAGTGTGTCGACGACAAATTCATATCCCTGA 50
864 uThrLeuMetSerMetProArgAlaPheIleArgLysValProGlySerV 881
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
51 GACCCCTGGTAACTTCCCGAAGTCCCTTATAGTCAAAAGTACCTGGTCCG 100
881 alValValThrIleCysThrSerGlyAlaSerAspArgLeuGluLeuArg 897
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
101 TCGTGGTTAGCATACACATTCGGGCATTTCCGACAAACTTGAATTCGG 150
898 GlyAlaPheAspLysSerLysGluThrPheGlyArgLysLeuLysAsnSe 914
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
151 GCGCGGTTCGACGTTCTCTAAAAAGAAATTTCCGAGGAGGTACGTTCCG 200
914 rArgLeuArgValPheSerArgAlaIleValGluAspSerIleLysValM 931
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
201 TCGTTCGGCGTATTCTTAGGCTATTGTGGAGGATACCATCAAGGTTA 250
931 etLysAlaMetLysThrGluAspGlyLysProLeuProIleThrGluAsp 947
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
251 TGAAGGCGATGAATCAGAGGATGTTAAACCACTCCCTATATGCGGAGAT 300
948 SerValTyrAlaPheIleMetGlyAsnValSerAsnValHisCysThrAr 964
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
301 TCGGTGTACGCGTTCATGACGAGCAATATGCAAACTTCATTCGACTAG 350
964 qAlaGlyLeuLeuGlyGlySerLysAlaThrValValSerSerValSerL 981
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
351 GGCTGGTTTCTCGCGGGGCTCAAGGGCTCGCGGGCTCTTTAGCTGTGA 400
981 ySGlyLeuValAlaAArgGlyAlaAlaThrLysAlaPheSerGlyIleThr 997
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
401 AGSGTGCAGCTTCACGCGCTACTGGAAACAACACTCTTTTCAGGTCTCACA 450

998 SerPhePheSerThrGlySerLeuPheTyrAspArgGlyLeuThrGluAs 1014
1014 pGluArgLeuAspAlaLeuValArgThrGluAsnAlaIleAsnSerProV 1031
501 AGAGAGCGTGTGATGACATACCGCGCGTGAACATCTGTGAATCACCTG 550
1031 aGlyIleLeuGluThrSerArgValAlaValSerLysValValAlaGly 1047
551 TAGGCGCTCTAGAACCTCGAGCTTCGGTTGGGAAGCGGGTCTTCGGA 600
1048 ThrLysGluPheTrpSerGluValSerLeuAsnAspPheThrThrPheVa 1064
601 ACGAAAGCTTTTCTCAGAAATGTTCATTTGGAGGACTTCACCACTTCGT 650
1064 lLeuArgAsnLysValLeuIleGlyIlePheValAlaSerLeuGlyAlaA 1081
651 CATAAAAATAGGGTGCATTATGGTGTATTTACTCTTTCCATGGCTCTCA 700
1081 laProIleAlaTrpLysTyrArgArgGlyIleAlaAlaAsnAlaArgArg 1097
701 CTCGGGTGCTGGAAGTACAGAAGGAATATCGCGCAACTGGCGTGGAT 750
1098 TyrAlaGlySerSerTyrGluThrLeuSerSerLeuSerSerGlnAlaAl 1114
751 GTTTTCCACCGTGCCTCGTTCGGGTACCGCGCCATCGGTTTACAAATGTCT 800
1114 aGlyGlyLeuArgGlyLeuThrSerSerThrValSerGlyGlySerLeuV 1131
801 TAGTCGGGAAGGTCGTAGCTGTGGTGGAGCGTCTGCGCGCTTAAACAG 850
1131 alValArgArgGlyPheSerSerAlaValThrValThrArgAlaThrVal 1147
851 TGACTCGAGGAGGCTATCTTCGGCGGTTCGGGTGACCAGAAATACAGTG 900
1148 AlaLysArgGlnValProLeuAlaLeuLeuSerPheSerThrSerTyrAl 1164
901 GCTAGCGCTCAGGTACCATTCGCGTGTGCTTCGTTTCCACGCTCTTAGCG 950
1164 aIleSerGlyCysSerMetLeuGlyIleTrpAlaHisAlaLeuProArgH 1181
951 ACTCAGTGGTTCGACTTTGTAGGTATTGTGGCTCATGCTCTCCCTAGGC 1000
1181 lsLeuMetPhePhePheGlyLeuGlyThrLeuLeuGlyAlaArgAlaSer 1197
1001 ATTTGATGCTTCTTCTTGCCCTAGGAGCGCTCTTCGGGTGAGTCCAGT 1050
1198 AlaAsnThrTrpLysPheGlyCysSerAsnAsnTrpCysAlaValPr 1214
1051 ACCAATTCGTGCTTGGGGCTATACGAACAGTCTGTTCAACCGTACC 1100
1214 oGluValValTrpArgGlyLysSerValSerSerLeuLeuLeuProIleT 1231
1101 GGAATTAACCTGGGAAGGAGGAGTTACAGATCTTTATTGCCCAAGCAG 1150
1231 hrLeuGlyValSerLeuIleIleArgGlyLeuLeuAsnAspThrIlePro 1247
1151 CTTTAGGTATTCTCTCGTGTGGCGGGTGTAAAGTGAACATGTGCCA 1200
1248 GlnLeuAlaTyrValProProValGluGlyArgAsnValTyrAspGluTh 1264
1201 CAACATACGTACGTACCCCGCATTCAGGTGGAATGTTTATGATCAGGC 1250
1264 rLeuArgTyrTyrArgAspPheAspTyrAspGluGlyAlaGlyProSerG 1281
1251 ACTAAATTTTATCCGACTTTGACTATGACATGGTGCAGGCCCATCCG 1300
1281 lyThrGlnHisGluAlaValProGlyAspAspAsnAspGlySerThrSer 1297
1301 GGACGGCTGGTCAAGCGATCCTGSA...ACCAATACTTCGATACTTCT 1347
1298 SerValSerSerTyrAspValValThrAsnValArgAspValGlyIleSe 1314

1348 TCGGTTTCTCT...GACGATGGTTGCC 1373
1314 rThrAsnGly...GluValThrGlyGluG 1323
1374 CGCTAGTCGGTGGCTTCGACCGCGGTGAGCAGGTCCACGCCATG 1423
1323 luGluThrHisSerProArg...SerValGlnTyrThrTyrValGluGlu 1338
1424 CTGTTGATGAATCACCAAGGGTAGTTGAGTTCGCTACAGAGAAGCT 1473
1339 ...GluValAlaProSerAlaValAlaGluArgGlnGlyAsp... 1352
1474 GTAGATGAACATCCGCGGTGTGAAGCTGAAGTTGAAAGGATCTAAT 1523
1353 ...ProSerGlySerGlyThrAlaAspAla...X 1362
1524 AACACCACTTGTACAGCTGTAGAGTCGCCCGCTAGGTCTCTGAAG 1573
1362 etAlaPheValGluSerValLysLysGlyValAspValPheHisGln 1378
1574 CTGGGAGCGCGCCCAACGTCGAGGAGGTGTGTCGGAGGTT... 1614
1379 GlnSerSerGlyGluThrAlaArgGluValGluValAspGlyLysGlyLe 1395
1615GAAGCTGAGAAATGTTCCGAGGTTCATCGTTGAC..... 1647
1395 uLeuProGluSerValValGlyGluAlaProThrGlnGluArgGlyArgA 1412
1648 .GTTCTAGTTCA.....GAACCGCGGTACAGAAGTCCCTGAAT 1687
1412 laAlaAspGlyAsnThrAlaGlnThrAlaValAsnGluGlyAspArgGlu 1428
1688 CAACCAATGGT.....GTCCAAGCTGCAAGAACTGAA.....GAG 1722
1429 ProValGlnSerSerLeuValSerSerProGlnAlaAspIleProLysVa 1445
1723 GTTGTGCAGGCGCACATGTGGAGCT.....GGGT 1754
1445 lThrGlnSerGluValHisAlaGlnLysGluValLysGlnGluValPro. 1461
1755 AGCTAAATCAGAAGTG...AGTCAACGTGTGTTCTCGCGAAGTACCG 1801
1462LeuAlaThrValSerGlyAla..... 1468
1802 CACATGAAGCTGGTCTTGAGGCATCTAGTGGCGCGTCTGGAGCCATTG 1851
1469ThrProIleValAspGluLysProAlaProSerVal...Th 1481
1852 CAAGTTTCTGTGCCAGTAGCCCTAGAGAAACTGTTTATCTGTCGAGAA 1901
1481 rThrArgGlyValLysIleIleAspLysGlyLysAlaValAlaHisValA 1498
1902 GCGCGCTCAGCTAAAGCGGTAGATAAGGCGGTAAGCGGCTGTCGACGCAA 1951
1498 laGluLysLysGlnValGlnValGluGlnProLysGlnArgSerLeuThr 1514
1952 AGGAAGTCAAGAAATGTACCGGTAAAGACGTATACCACAGGGGCTCTAAA 2001
1515 lIleAsnGluGlyLysAlaGlyLysGlnLeuCysMetPheArgThrCysSe 1531
2002 ATTAGTAGGATACCGTTCGTAAGGAATGTGCATGTTTAGAACGTGTC 2051
1531 rCysGlyValGlnLeuAspValTyrAsnGluAlaThrIleAlaThrArgP 1548
2052 CTGCGCGCTGCAGTTGGAGCTGTACAACTGAACGACCATCGCCACTAGT 2101
1548 heSerAsnAlaPheThrPheValAspAsnLeuLysGlyArgSerAlaVal 1564
2102 TCTCAATCGCTTTTACCTTTGTCGATAGCTTGAAAGGAGGAGTCCGCTC 2151
1565 PhePheSerLysLeuGlyGlyTyrThrTyrAsnGlyGlySerHisVa 1581
|||||

2152 TTTTCTCAAAGCTGGGTGAGGGGTATACCTATAATGGTGGTAGCCATGT 2201
1581 lSerSerGlyTyrProArgAlaLeuGluAspIleLeuThrAlaIleLysT 1598
2202 TTCATCAGGGTGGCCTCGGCCCTAGAGATATCTTAACGGCAATTAAGT 2251
1598 yCProSerValPheAspHisCysLeuValGlnLysTyrLysMetGlyGly 1614
2252 ACCCAAGCGCTTCGACCACCTGTTAGTGCAGAAGTACAGAATGGGTGA 2301
1615 GlyValProPheHisAlaAspGluGluCysTyrProSerAspAsnPr 1631
2302 GCGTACCATTCCACGCTCATCAGGAGGAGTGCTATCCATCAGATAACCC 2351
1631 oIleLeuThrValAsnLeuValGlyLysAlaAsnPheSerThrLysCysA 1648
2352 TATCTTGACGGTCAATCTCGTGGGAAGCAAACTTCGACTAAGTGCA 2401
1648 rGlySGlyGlyLysValMetValIleAsnValAlaSerGlyAspTyrPhe 1664
2402 GGAAGGGTGGTAAGGTCAATGGTCATAAACGTAGCTTCGGGTGACTATTTT 2451
1665 LeuMetProCysGlyPheGlnArgThrHisLeuHisSerValAsnSerIl 1681
2452 CTTATGCCCTTCGGGTTTCAAAGGACGCACTTGCAATTCAGTAACCTCAT 2501
1681 eAspGluGlyArgIleSerLeuThrPheArgAlaThrArgValPheG 1698
2502 CGACCAAGGGCGCATCAGTTTGACGTTTCAGGCAACTCGGCGCGCTTTTG 2551
1698 lYValGlyArgMetLeuGlnLeuAlaGlyValSerAspGluLysSer 1714
2552 GTGTAGCGAGGATGTTCAGTATAGCGCGCGGTGTCGGATGAGAAGTCA 2601
1715 ProGlyValProAsnGlnGlnProGlnSerGlnGlyAlaThrArgThrIl 1731
2602 CCAGGTGTTCCAAACCAAGCAACACAGACCAAGGTGCTACCAGAACAT 2651
1731 eThrProLysSerGlyGlyLysAlaLeuSerGluGlySerGlyArgGluV 1748
2652 CACACCAAAATCGGGGGCAAGGCTCTATCTGAGGGAAGTGTAGGGAAG 2701
1748 aLYSGlyArgSerThrTyrSerIleTrpCysGluClnAspTyrValArg 1764
2702 TCAAGGGGAGGTGCGACATACTCGATATGTGCGAACAAAGATTACGTTAGG 2751
1765 LysCysGluTrpLeuArgAlaAspAsnProValMetAlaLeuGluProAs 1781
2752 AAGTGTAGTGGCTCAGGGCTGATATCCAGTGATGGCTCTTAACCTCG 2801
1781 pTyrThrProMetThrPheGluValValLysThrGlyThrSerGluAspa 1798
2802 CTACACCCCAATGACATTTGAAGTGGTTAAAGCCGGGACCTCTGAAGATG 2851
1798 lAValValGluTyrLeuLysTyrLeuAlaIleGlyIleGluArgThrTyr 1814
2852 CCGTCGTGGAGTACTTGAAGTATCTGGCTATAGGCATTCGGAGGACATAC 2901
1815 ArgAlaLeuLeuMetAlaArgAsnIleAlaValThrAlaGluGlyVa 1831
2902 AGGGCGTTGCTTATGGCTAGAATATTCGCCGTCACCTACCCCGCAAGGTGT 2951
1831 lLeuLysValProAsnGlnValTyrGluSerLeuProGlyPheHisValT 1848
2952 TCTGAAGGTACCTAATCAAGTTTATGAATCACTACCGGGCTTTCAGTTT 3001
1848 yLysSerGlyThrAspLeuIlePheHisSerThrClnAspGlyLeuArg 1864
3002 ACAAGTCGGGCACAGATCTCATTTTTCATTCACACACAGACGGCTTGCGT 3051
1865 ValArgAspLeuProTyrValLeuIleAlaGluLysGlyIlePheThrLy 1881
3052 GTGAGAGACCTACCGTACGTATTCATAGCTGAGAAGGTATTTTATCAA 3101

1881 sGlyLysAspValAspAlaValValAlaLeuGlyAspAsnLeuPheValIC 1898
3102 GGCAAAGATGTGACGCGGTAGTAGCTTTGGCGACAAATCTGTCGGTAT 3151
1898 yAspAspIleLeuValPheHisAspAlaIleAsnLeuIleGlyAlaLeu 1914
3152 GTGATGATATATTGGTTTCCATGATGCTATTAAATTTGATGGGTGCATC 3201
1915 LysValAlaArgCysGlyMetValGlyGluSerPheLysSerPheGluTy 1931
3202 AAAGTTGCTCGATGTGTTATGGTGGGTGAATCATTTTAAGTCTTCGAATA 3251
1931 rLysCysTyrAsnAlaProProGlyGlyGlyLysThrThrLeuValA 1948
3252 CAAATGCTATAATGTCTCCCGAGGTGGCGTAAAGACGATGCTAGTGG 3301
1948 sPGLuPheValLysSerProAsnSerThrAlaThrIleThrAlaAsnVal 1964
3302 AGAATTTGTCAAGTCACCCATAGCACCGCCACCATTTACGGCTAACGTG 3351
1965 GlySerSerGluAspIleAsnMetAlaValLysLysArgAspProAsnLe 1981
3352 GGAAGTTCTCAGGACATAAATATGGCGGTCAAGAAGAGAGATCCGAATTT 3401
1981 uGluGlyLeuAsnSerAlaThrThrValAsnSerArgValValAsnPheI 1998
3402 GGAAGGTCTCAACAGTGTACCACAGTTAACTCCAGGGTGGTTAACTTTA 3451
1998 lEValArgGlyMetTyrLysArgValLeuValAspGluValHisMetMet 2014
3452 TTGCTAGGGGAATGTATAAAGGGTTTTGCTGGATGAGGTGTACATGATG 3501
2015 HisGlnGlyLeuLeuGlnLeuGlyValPheAlaThrGlyAlaSerGluGl 2031
3502 CATCAAGGCTTACTACAACTAGCGGTCTTCGCAACCGCGCTCGGAAGG 3551
2031 yLeuPhePheGlyAspIleAsnGlnIleProPheIleAsnArgGluLysV 2048
3552 CCTCTTTTGGAGACATAAATCAGATACCATTCATATAAACCGGAGAAAG 3601
2048 aPheArgMetAspCysAlaValPheValProLysLysGluSerValVal 2064
3602 TGTATTAGGATGGATTGTCTGTATTGTTCCAAAAGAAAGGAAAGCGTTGTA 3651
2065 TyrThrSerLysSerTyrArgCysProLeuAspValCysTyrLeuLeuSe 2081
3652 TACACTTCTAAATCATACAGGTGTCGGTTAGATGTTTGTACTTTGTGTGC 3701
2081 rSerMetThrValArgGlyThrGluLysCysTyrProGluLysValValS 2098
3702 CTCAATGACCGTAAGGGGAACGAAAGTGTACCCTGAAAAGGTGCTTA 3751
2098 eGlyLysAspLysProValValArgSerLeuSerLysArgProIleGly 2114
3752 GCGGTAAAGACAAACCAAGTAGTAAGATTCGTCCTCAAAAAGGCCAATTGA 3801
2115 ThrThrAspAspValAlaGluIleAsnAlaAspValTyrLeuCysMetTh 2131
3802 ACCACTCATCAGCTACGCTCAAAATAAACGCTGACGTGTACTTTGTCATGAC 3851
2131 rGlnLeuGluLysSerAspMetLysArgSerLeuLysGlyLysGlyLysG 2148
3852 CCAGTTGGAGAAGTCGGATATGAAGAGGTGCTTGAAGGGGAAAAGGAAAG 3901
2148 luThrProValMetThrValHisGluAlaGlnGlyLysThrPheSerAsp 2164
3902 AAACACCAAGTGTGACAGTGCATGAAGACACAGGGAACAAATTCAGTGT 3951
2165 ValValLeuPheArgThrLysLysAlaAspAspSerLeuPheThrLysGl 2181
3952 GTGGTATTGTTTAGGACGAAGAAAGCCGATGACTCCCTATTCTACTAAACA 4001

```
2181 nProHisIleLeuValGlyLeuSerArgHisThrArgSerLeuValTyrA 2198
|||||
4002 ACCGCATATACTTGTGTTGTCGACACACACACACGCTCACTGTTTATG 4051
|||||
2198 IaAaLeuSerSerLysLeuAspLysValGlyThrTyrIleSerAsp 2214
|||||
4052 CGGCTCTGAGCTCAGATGGACGATAGGTCGGCACATATATTAGCGAC 4101
|||||
2215 AlaSerProGlnSerValSerAspAlaLeuLeuHisThrPheAlaProAl 2231
|||||
4102 CGGTCGCGCTCAATCAGTATCCGACGCTTGGCTTCACACGTCGCCCGGC 4151
|||||
2231 aGlyCysPheArgGlyIle 2237
|||||
4152 TGGTTGCTTTCAGGTATA 4170
|||||

seq_name: gb_vil:AF037268

seq_documentation_block:
LOCUS AF037268 13154 bp RNA VRL 28-APR-1998
DEFINITION Grapevine leafroll-associated virus 3, partial genomic sequence.
ACCESSION AF037268 U82937
VERSION AF037268.1 GI:3090872
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
COMMENT
FEATURES
source
1. .13154
/organism="grapevine leafroll-associated virus 3"
/strain="NY1"
/db_xref="taxon:55951"
<1. .2107
/function="replication and unknown functions"
/note="incomplete ORF1a, one-third of the expected ORF1a;
sequenced portion contains helicase domains identified by
sequence comparison"
/codon_start=2
/product="unknown"
/protein_id="AAC40717.1"
/db_xref="GI:3090885"
/translation="DVYNEATATRESNAFTFDNLKGRSAVFESKLGEGTYNGGSH
VSSGWPRALDILTAIKYPSVEDHCLVQYKMGGVPPFIADDEECPSPNLTIVNLV
GFVGRMLQAGVDEKSGVPNPQQSQGATRIITPKSGKALSSEGSGREVKGRTY
STWCEQDYVRKCEWLADPNVMALEPDYTPMTPEVVKGTSEDVAVVEYLKLAIGIER
TVRALMARNIAVTTAEGVLKVPNOVYESLPGPHYKSGTDLIHFHSTQDGLVRDLPY
VLAIEGKIFTKGKDVAVALGDNLFVCDIILVFHDAINLIGALKVARGCMGESPKS
FEYKCNAPPGGGKTTLVDEFPVKNSSTATITANVGSSEDINMAYKKRDPNLEGLNS
ATTVNSRVNFIVRGKIKRVLVDEVHHMHQGLQLQGVFAFGASEGLFFGDIQIPIN
```

```
REKVRMDCAVFPKKESVYVYTSKSRCPDVCYLLSSMTVRTEKCYPKVYSGDKK
PVYRSKRPITGTTDDVAEINADVYLCMTQLEKSDMKRSLKGGKETPVMTVIEAGK
TFSDVLFRTKKADDSLTQKPHILVGLSRHTRSLVVAALSSKLLDDKVGTYISDAPSQ
SVSDALLHTFAFACGFRGI"
2112..3713
/function="replication"
/note="ORF1b; probably expressed via +1 frameshift
mechanism; 61 kDa RNA-dependent RNA polymerase"
/codon_start=1
/product="RNA-dependent RNA polymerase"
/protein_id="AAC40705.1"
/db_xref="GI:3090873"
/translation="MNFPGTFEGELVRKIPTSHFVAVNGFELEDLDGCPAFDFFED
DFETSDQSELIEDVRISESHEFTSKIEDRFYSIRSSVGLPKRNTLKCLNLTFFENRN
FNADRCNVGCDSDVAHELKEIFEEVNVNKAARAEVTESHLSNTMLLSDLMTKRAPN
AYSKRALGSEFVHPMLTSTYLMVKADVKPKLDNTPLSKYVTGQNIYVHDCVTAL
FSCITACVERLKYVVDREKLFYHGMDTAEALARNLNLGDIROYTYIELDISKYDKS
QSMKQVEELIULTLGVREVLSTFFCGEYDSVVRMTKELVLSVGSQRRSGGANTW
LGNVLCTLLSVLRGLDYSYIVSGDDSLIFSRQPLDIDTSLVDSNFGFDVKIFNQ
AAPYFCSKFLVQVEDSLFFVPDPLKLFVKFGASKTSDIDLHIEIFQSFVDSLKGFRRE
DVIQELAKLVTRKYKHSGWTYSALCVLHVLNSANFSQFCRLYYHNSVNLVDVRPIQRTES
LSLALAKARILRWKASRAFAFSIKRG"
3943..4098
/function="unknown"
/note="ORF2; seems to have no homologs in genomes of other
closteroviruses"
/codon_start=1
/product="6 kDa protein"
/protein_id="AAC40706.1"
/db_xref="GI:3090874"
/translation="MYSRGSFKSRVLTPLTVGAYMWEFELPYLTKRHSYSAPSVA
TFSLSVR"
5165..5302
/function="unknown"
/note="ORF3; similar to small hydrophobic proteins of
other closteroviruses"
/codon_start=1
/product="5 kDa protein"
/protein_id="AAC40707.1"
/db_xref="GI:3090875"
/translation="MDDFKQAILLVVDFVFVILLVLTFFVPRLQSSSTINVTGLRT
V
5321..6970
/function="molecular chaperone"
/note="ORF4; identified by sequence comparisons; similar
to HSP70"
/codon_start=1
/product="59 kDa protein"
/protein_id="AAC40708.1"
/db_xref="GI:3090876"
/translation="MEVGIDFGTFTICFSPSGVSGCTPVAGSVVYVETQIFPEGSS
TYLIGKAAGKAYRDGVEGRLYVNPKRWAGVTRDVERVYVKLPKTYTVKIDSGALLI
GGLGSGPDTLLRVVDVICLFLRALILECERYTSTTAAVTVPADYNSPKRSFVVEA
LKGLGIPRGVYVNEPTAAALYSLAKSRVEDLLLVDFDGGGTEDFVFKKGNILCVI
FSVGDNLGGRDIDRAIVEIKQIKRGASDAKLGIFVDSMKEDLSNNNAITQHLIPV
EGGVENVLDTSDELDAIVAPFASARAVEVFTGLDNFYPDPVIAVMTGGSSALVKVRSD
VANLPQISKVWFSDTDFRSCVACCAKVCYCDTLAGNSRLGLVDTLTNTLDEVVGLQPV
VIRPKGPSPCVSTHRYTVGGGVVYIFEGENNAFLNEPTFRGYSKRGDDVETDV
AQNLSVDGTVSVIVNGEEVKNELVPGTNNVLDLSLYKSGREDLEAKAIPVLTILN
ILHDKAFTRNLGNKDKGFSDLRIENFLKSAVDTDTILNG"
6963..8414
/function="unknown"
/note="ORF5; similar to genes found in a similar position
in other closteroviruses"
/codon_start=1
/product="55 kDa protein"
/protein_id="AAC40709.1"
/db_xref="GI:3090877"
/translation="MDKYIYVTGILNPNEARDEVFVYVNGYIGPGGRSPNRGSKYV
VYVWNSAARTISGFTSTOSTIDAFAYFLKGLGTLTNSPINCENWVRSSKDLASAPFR
TLIKGKIYASRSVDSNLPKDRDDIMEASPRLSPSDAAACRAVSOVGVYVDYTONLE
STIVPLRWWEIKARRGSAHSLPKVYSVYDFTYTNQLQELLSDEVTARTDTVSAYATD
SMAFLVRMULPTLRGQWLKDLVGLVLLVRRRPFANFSYDVRVAVWYDVIATLKLURLFF
```

```
CDS
NKDTPGGIKDLKPCVIESFDPFHELSVFSRLSYEMTTGKGKICPEIAEKLVRLM
ENYKLTPTWALIIILVYISVGTNATRIKRDPDFLNRKIKRVERKSLRGVEDRA
PRISEKGINAQRVLCRYSDLTCLARRHYGIRNNWKTLSYVDGTLAYDADCTISK
VNTINTADHASIIHYIKTNEQVVTGTLPHQL"
8504..9445
/note="ORF6"
/codon_start=1
/evidence=experimental
/product="35 kDa coat protein"
/protein_id="AAC40710.1"
/db_xref="GI:3090878"
/translation="MAFELKLGQIYEVVPPENNLRRVRVGDAAQGFKSKASFLKYVKDGT
QAEITGAVPEKVFATAALAAQEPPOPAOVAEPQETDIGVWPESETLTENKL
VFEKDPDLKMGKIALDLAGVTHKPKVINEKQSVSEVAMKINAALMELCKVMG
ADDAATKTEFLYVWQIACFTFTSSSEFEKEDIETDDGKKIYAVWYDICKQRAAS
TGYENPVKYLAYFTPTITATLNGKLVMNEKVMQAQHGVPKPFPPYITDCVRPTVDLF
NDAILLAAMRAQQAQRNKVTADNTLHNVLQLQK"
9508..10941
/note="ORF7: diverged copy of the GLraV-3 coat protein"
/codon_start=1
/product="53 kDa protein"
/protein_id="AAC40711.1"
/db_xref="GI:3090879"
/translation="MGAYTHVDFHESRLLDKQDYLSEKSADEAPPDPGVPVPSVY
RAYLQIADFPNTQSLSVTLISANKLASGLMGSDAVSSSPMLMNDVDGYFECGWCHN
KPYLGRVIFCRKVIIGRGVEITTKKNVTSNNWNEASVYIQVNVVDGLAQTVTSTYT
QTDVSGLPKNWTRIKYIKIVSDNLIYPCFSDSKLGMRIKRLSLVSPVIRFDIL
LKLKPSKARIEDVNDIDTSLVPSVPVPESTGGVPSQOLDVVALTSDVTELNT
RQOKICFPDSVLSINEADYIDERYLPITEALQINARLRVLVLSKGSQTPRDMGMI
VAMIQFLVYSTVKNISVKDGYRVETELQKRVLYSYSEVREALGGYKYGASPTNVR
SEMYFAHTTITLIEKKIOPACTALAKHGVKRPFTPCYCFDFALLDNRYYPADVLKAN
MACAIAKSNLRKSGSETYNILESI"
10952..11509
/function="unknown"
/note="ORF8: seems to have no homologs in other
closteroviruses"
/codon_start=1
/product="21 kDa protein"
/protein_id="AAC40712.1"
/db_xref="GI:3090880"
/translation="MEFRPLITVRRDGVNTGSLKVIAYDLHYDNIPDNCAVKSFRD
TDTGTVWKEYSTNSAFILSPYKLFSAVFNKEGEMISNDVGSSEFRVYNIFSOMCKDIN
EISEIQARGLYLETVLGGQADTDIFEDVLTNNKAVRLVKNKDHSAWCGILNDLKWE
SNKEKFGKRDILDTYVLSDDYPGFK"
11506..12039
/function="unknown"
/note="ORF9: seems to have no homologs in other
closteroviruses"
/codon_start=1
/product="19.6 kDa protein"
/protein_id="AAC40713.1"
/db_xref="GI:3090881"
/translation="MKLLSLRYLILRLSKSLRTNDHLVILIKALINYNASPTDEG
AVLDRSREINFLVARCGSONSCVMKALITNTVKMSIETARSFIDGLILVADSSV
SALPEAKSIKDNFRLKRGKYYSGDCGSDVAKVKYILSGENKGLGCVDSLKLVCG
RQGGNVQLHLILSLG"
12046..12585
/function="unknown"
/note="ORF10: seems to have no homologs in other
closteroviruses"
/codon_start=1
/product="19.7 kDa protein"
/protein_id="AAC40714.1"
/db_xref="GI:3090882"
/translation="MDLSFIIVQILSASYNNDVTALYTLINAYSDVDTTRWAAINDP
QAEVNVKAVVATTATTELHETLILIDSISAFAYDVOVGLVGIARGLRHSEDLVEI
KSMELFVCRKRSKRYLGLYLSDOCTNKYMLIQAGLAUAVEGADILKTNHLVSGNKF
SPNFGIARMLLLTITCCGAL"
12588..12698
/function="unknown"
/note="ORF11: seems to have no homologs in other
closteroviruses"
/codon_start=1
```

```
alignment_scores:
  Quality: 3644.00      Length: 701
  Ratio: 5.198          Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-301-906-15 x AF037268 ..
Align seg 1/1 to: AF037268 from: 1 to: 13154

1537 AspValTyrAsnGluAlaThrIleAlaThrArgPheSerAsnAlaPheTh 1553
|||||
2 GATGTGTACACGAGGAGCATATGCCACCGAGGTCTCAACAGCATTTAC 51
|||||

1553 rPheValAspAsnLeuLysGlyArgSerAlaValPhePheSerLysLeuG 1570
|||||
52 CTTTGTGCATAACTTGAAGGAGGAGTGGCGTCTTTTCTCAAGCTGG 101
|||||

1570 lyGluGlyTyrThrTyrAsnGlyGlySerHisValSerSerGlyTrpPro 1586
|||||
102 GTCAGGGGTATACCTATAATCGTGTGTAGCCATGTTTCATCAGGTGGCCT 151
|||||

1587 ArgAlaLeuGluAspIleLeuThrAlaIleLysTyrProSerValPheAs 1603
|||||
152 CGTGCCTTAGAGGATATCTTAACGGCAATTAACTACCAAGCGTCTTCGA 201
|||||

1603 phiScysLeuValGlnLysTyrLysMetGlyGlyGlyValProPheHisA 1620
|||||
202 CCACTGTTTAGTCAGAAAGTACAAGATGGGTGGAGGCGTACCATTCCACG 251
|||||

1620 laaspaspGluGluCysTyrProSerAspAsnProIleLeuThrValAsn 1636
|||||
252 CTGATGACGAGGAGTGTCTATCCATCAGATAACCTATCTTCACGGTCAAT 301
|||||

1637 LeuValGlyLysAlaAsnPheSerThrLysCysArgLysGlyGlyLysVa 1653
|||||
302 CTCGTGGGGAAGGCAAACTTCTCGACTAAGTGCAGGAAGGTGTGAAGT 351
|||||

1653 lMetValIleAsnValAlaSerGlyAspTyrPheLeuMetProCysGlyP 1670
|||||
352 CATGGTCATANAACGTAGCTCGGGTGACTATTTCTTATGCTTGGCGGTT 401
|||||

1670 heGlnArgThrHisLeuHisSerValAsnSerIleAspGluGlyArgIle 1686
|||||
402 TTCAAAGGACGCACTTGCAATTCAGTAAACTCCATCGACGAAGGCGCATC 451
|||||

1687 SerLeuThrPheArgAlaThrArgArgValPheGlyValGlyArgMetLe 1703
|||||
452 AGTTTGACGTTTCAGGGCAACTCGGCGGTCTTTGGTGTAGGAGGATGTT 501
|||||

1703 uGlnLeuAlaGlyValSerAspGluLysSerProGlyValProAsnG 1720
|||||
502 GCAGTTAGCCGGCGCGGTGTCGATGAGAGTACCAGGTGTTCACAAACC 551
|||||

1720 lGlnProGlnSerGlnGlyAlaThrArgThrIleThrProLysSerGly 1736
|||||
552 AGCAACCACAGAGCAAGGTGCTACCAGAAATCACACCAAAATCGGGG 601
|||||

1737 GlyLysAlaLeuSerGluGlySerGlyArgGluValLysGlyArgSerTh 1753
|||||
602 GGCAAGGCTCTATCTGAGGGAAGTGGTAGGGAAGTCAAGGGGAGGTGCAC 651
|||||

1753 rTyrSerIleTrpCysGluGlnAspTyrValArgLysCysGluTrpLeuA 1770
|||||
652 ATACTCGATATGTGCGCAACAGATTACGTAGGAAGTGTGAGTGGCTCA 701
|||||

1770 rGalaAspAsnProValMetAlaLeuGluProAspTyrThrProMetThr 1786
|||||
702 GGCTGATAATCCAGTGTATGGCTCTTCAACCTGACTACACCCCAATGACA 751
|||||

1787 PheGluValValLysThrGlyThrSerGluAspAlaValGluTyrLe 1803
|||||
```

```
752 TTTGAAGTGGTTAAACCGGGACCTCTCTGAAGATGCCGTCGTCGAGTACTT 801
1803 ubystyrLeuAlaIleGlyIleGluArgThrTyrArgAlaLeuLeuMeta 1820
|||||
802 GAAGTATCTGGCTATAGCATTGACAGGACATACAGGGCGTTGCTATGG 851
1820 IaArgAsnIleAlaValThrThrAlaGluGlyValLeuLysValProAsn 1836
|||||
852 CTAGAAATATATGCCGCTACTACCGCGAAGGTGTTCTGAAAGTACCTAAT 901
1837 GlnValTyrGluSerLeuProGlyPheHisValTyrLysSerGlyThrAs 1853
|||||
902 CAAGTTTATGAATCACTACCGGGCTTTTCCAGTTTACAAGTCGGGCACAGA 951
1853 pLeuIlePheHisSerThrGlnAspGlyLeuArgValArgAspLeuProT 1870
|||||
952 TCTCATTTTTCATTCAACACAGACGGCTTGGCTGTGAGACCTACCGT 1001
1870 yrValLeuIleAlaGluLysGlyIlePheThrLysGlyLysAspValAsp 1886
|||||
1002 ACGTACTCATAGCTGAAAAGGTATCTTTACCAAGGGCAAGATGTCGAC 1051
1887 AlaValValAlaLeuGlyAspAsnLeuPheValCysAspAspIleLeuVa 1903
|||||
1052 GCGGTGGTAGCTTTGGGGACAATCTGTTCGTATCGGACGATATATCGT 1101
1903 lPheHisAspAlaIleAsnLeuIleGlyAlaLeuLysValAlaArgCysG 1920
|||||
1102 TTTCCACGATGCCATTATTTGATAGGTGCACCTGAAAGTCGCTCGATCGG 1151
1920 lymetValIglyGluSerPheLysSerPheGluTyrLysCysTyrAsnAla 1936
|||||
1152 GCATGGTGGCGAATCGTTTAAGTCCTTCGAATATAAGTGTATATGCT 1201
1937 ProProGlyGlyLysThrThrThrLeuValValAspGluPheValLysSe 1953
|||||
1202 CCCCAGGTGGCGGTAAAGACGACGAGCTTAGTGGACGAATTCGTTAAGTC 1251
1953 rProAsnSerThrAlaThrIleThrAlaAsnValGlySerSerGluAspI 1970
|||||
1252 ACCCAATAGCAGACCCATTACGGCTAATGTGGAAAGTTCTGAGGACA 1301
1970 IaAsnMetaIalavallysylsArgAspProAsnLeuGluGlyLeuAsnSer 1986
|||||
1302 TAAATATGCGGTGAAGAAGAGAGATCGGAATTTGGAAGGTCTCAACAGT 1351
1987 AlaThrThrValAsnSerArgValValAsnPheIleValArgGlyMetTy 2003
|||||
1352 CTTACACAGTTTAACCTCCAGGTGGTAAACTTTATCGTCAGGGGAATGTA 1401
2003 rLysArgValLeuValAspGluValHisMetMetHisGlnGlyLeuLeuG 2020
|||||
1402 TAAAGGGCTTTTGGTGGATAGGTGCACATGATCATCAAGCCTTACTAC 1451
2020 lNleuGlyValIpheAlaThrGlyAlaSerGluGlyLeuPhePheGlyAsp 2036
|||||
1452 AACTAGGGGTCTTCGAACCGCGCGGTGCGAAGGCCCTCTTTTGGAGAC 1501
2037 lIleAsnGlnIleProPheIleAsnArgGluLysValPheArgMetAspCy 2053
|||||
1502 ATAAATACATACCACTTATCAACAGGAGAGAGGTGTTTAGATGGATG 1551
2053 sAlaValPheValProLysLysGluSerValValTyrThrSerLysSerT 2070
|||||
1552 TGTCTTTTGTTCACAAAGAGGAAGCGTTGTATACACTTCTTAATCGT 1601
2070 yrArgCysProLeuAspValCysTyrLeuLeuSerSerMetThrValArg 2086
|||||
1602 ACAGGTGTCGCTAGATGTTTGTCTACTACTTGTCTCCTCAATGACCGTAAGG 1651
2087 GlyThrGluLysCysTyrProGluLysValValSerGlyLysAspLysPr 2103
|||||
1652 GAAACGGAAAAGTGTACCCCTGAAAAGGTCTGTTAGCGGTAGGACAAACC 1701
```

```
2103 oValValArgSerLeuSerLysArgProIleGlyThrThrAspVala 2120
|||||
1702 AGTAGTAAGATCGCTGCCAAAAGCCCAATTGGAACCACTGATGACGTAG 1751
2120 IaGluIleAsnAlaAspValTyrLeuCysMetThrGlnLeuGluLysSer 2136
|||||
1752 CTGAATAAACCGCTGACGTGTACTTTGTCATGACCCAGTTGGAGAAGTCC 1801
2137 AspMetLysArgSerLeuLysGlyLysGlyLysGluThrProValMetTh 2153
|||||
1802 GATATGAAGAGGTCTGTGAAGGGAAAGAAACACACACAGTGTATGAC 1851
2153 rValHisGluAlaGlnGlyLysThrPheSerAspValValLeuPheArgT 2170
|||||
1852 AGTGCATGAGCACAGGGAACATTCAGTCATGTGGTATTGTTTAGGA 1901
2170 hrLysLysAlaAspSerLeuPheThrLysGlnProHisIleLeuVal 2186
|||||
1902 CGAAGAAACCGCATGACTCCCTATTCTACTAAACACACCGCATATACTTGT 1951
2187 GlyLeuSerArgHisThrArgSerLeuValTyrAlaAlaLeuSerSerly 2203
|||||
1952 GGTTTGTGAGACACACACGCTCACTGGTTTATGCCGCTCTCGAGCTCAA 2001
2203 sLeuAspAspLysValGlyThrTyrIleSerAspAlaSerProGlnSerV 2220
|||||
2002 GTTGACCATAGGTGCGGCACATATATTAGCAGCGCTCACCTCAATCAG 2051
2220 aISerAspAlaLeuLeuHisThrPheAlaProAlaGlyCysPheArgGly 2236
|||||
2052 TATCCGAGCTTTGCTTCACAGCTTCGCCCGCGCTGGTTCGTTTCGAGGT 2101
2237 Ile 2237
|||
2102 ATA 2104
seq_name: gb_vii:AF283103
seq_documentation_block:
LOCUS AF283103 14861 bp DNA VRL 17-SEP-2000
DEFINITION Pineapple mealybug wilt associated virus-2 polyprotein (ORF1a) and
RNA-dependent RNA polymerase genes, partial cds; and hydrophobic
protein p5, heat shock protein 70, p46, coat protein, diverged coat
protein, p20, p22, and p6 genes, complete cds.
ACCESSION AF283103
VERSION AF283103.1 GI:10179918
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Melzer,M.J., Karasev,A.V., Sether,D.M. and Hu,J.S.
TITLE Nucleotide sequence, genome organization, and phylogenetic analysis
of Pineapple mealybug wilt associated virus-2
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 14861)
TITLE Melzer,M.J., Karasev,A.V., Sether,D.M. and Hu,J.S.
JOURNAL Direct Submission
Submitted (28-JUN-2000) Plant Pathology, University of Hawaii, 3190
Maile Way, Honolulu, HI 96822, USA
FEATURES
source 1..14861
/db_xref="taxon:136234"
gene <1..5520
<1..5520
CDS <1..5520
/gene="ORF1a"
/codon_start=1
/product="polyprotein"
/protein_id="AAG13938.1"
```


alignment_scores:

Quality: 2333.00 Length: 2251
Ratio: 1.927 Gaps: 55
Percent Similarity: 53.798 Percent Identity: 29.409

alignment_block:

US-09-301-906-15 x AF283103 ..

Align seg 1/1 to: AF283103 from: 1 to: 14861

```
103 AlaArgThrSerThrSerAlaThrSerAspValSerAspPheAspValVa 119
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
28 GCCCGCACTGACGCTCTTGTGTA...GACGTGCTCGATTAGATTATA 74
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
119 lPheGluAlaValSerAsnAlaLeuLeuValValHisTyr..... 132
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
75 CCACCTTAGGCTATCCCGCAGTCGCCCTGGTGGTAGTATATATTCGTCAC 124
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
133 .....HisArgValValProTyrAlaProVal 141
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
125 GGGTTGCAACGAATAACTTCATCAAAAGCCCGCACCTGTTAAATCGAG 174
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
142 LysArgGluInProLysProAlaValLysGlnAspGluGlnLysProLy 158
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
175 AGGCGTAGTCTCCACGGGTACGGGTGCCAGCAGATCAGCAGGCTAT 224
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
158 sArgGlnAlaSerHisTrpAlaValLysProThrAlaValGlyValHisv 175
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
225 ACCCGATGCC...CGATGG..... 240
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
175 alProLeuProLysLysGlnGluAlaLeuGlu.....ProAlaGlnSer 189
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
241 .....AAAAAGTACGAAGTTATGGATGGGCTCGATCGTAGAGCT 279
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
190 valProGlnInSerLeuGluCluLysAlaAlaLeuThrPheGlyLeuPh 206
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
280 TGTCT.....CGGACTCTGACG..... 297
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
206 ePheSerLysGlyGlyAspGluSerAspAlaValIleLeuArgLysG 223
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
298 ....ACCGGTGGCCCGTCGACAAGTTGACGGGTTGGTTATTACTAATG 343
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
223 lYLeuLeuPheAsnArgAlaLeuAsnValProIleAspValLysAsnThr 239
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
344 TCAAACTGTACTCG..... 357
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
240 PheValTrpAlaLysIleTrpAspGluAlaSerArgArgGlyTyrPh 256
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
358 .....TGGATACGTATGTACAGTCATTACTTGAACCGGTACACGTGGCT 401
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
256 cTyrValLysAspArgAlaValLysPhePheProIleValArgGlyArg 273
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
402 GTACTTCAGGCCCGGGGTGTTCAGATGCAGTTTGCATCCGATACC 451
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
273 laThrIleGluAspPheIleValAsnThrAlaProGlyCysAspValAla 289
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
452 CCACCGTCGCGGATTTTGTGTGACGCGGTGTAAGACAGCATGCGCGCG 501
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
290 LeuProArgIleGluLeuTrpSerMetArgGluArgAlaPheValCysTh 306
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
502 CTGCGTAAACTGCGCTTTATTTGCCCGGTTCTTCACAGGAGCGCTTGAG 551
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
306 rThrLysGlyTrpCysTrpPheAsnAsnGluArgLeuArgGlyGluIleT 323
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
552 CTCTCGAGGGTGGTGGTGTGTTAAACCATACCGCGGTGCAGAAAGCTTTAA 601
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
323 yrArgArgArgCysPheSerSerPheSerIleGlyPheLeuMethHis 339
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
602 CACACTGTGAACATTTTCCGAGCATAGTCTCTATTTCATACCTGAGGAAG 651
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
340 LeuGlyPheArgSerLeuLysValIleArgPheAlaGlyThrAsnIleLe 356
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
652 CTCGGGTGTGGCGGCTTCAAGGTTATGCCACAGGGGATCCCAAGGTGA 701
```

```
356 uHisMetProSerLeuAsnGluGluArgThrPheGlyTrpLysGlyAla 373
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
702 CCATTTTTCGAATAATAAT.....G 721
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
373 spValTrpLeuProAsnValProLysThrAlaIleValAlaGlyAspArg 389
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
722 GTATACTTATAGATAGCCTCTCATATCACTTCGCTGGCGGTGAGAGG 771
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
390 ThrArgLeuGlyGlyIleLeuAlaSerValAlaAsnAlaLeuAsnGl 406
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
772 GCCCGAGTGGTACAGAGCTGTAGCATCTATCGCAAGCGCTCAACCA 821
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
406 nGluGluValTyrSerValValSerIleThrAsnArgLeuValL 423
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
822 GGAAGAGGTCTACGCTGGAGTTGTCTCATCCATACTCAAAAGTTAATGC 871
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
423 euArgAspGlnSerAlaLeuSerHisLeuAspThrLysLeuCysAsp 439
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
872 TGAAGAGACCAAGCGCTGTGTTAATACATTGATACCCAGCTCTACAAG 921
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
440 MetPheSerGlnArgAspAlaMetIleArgGluLysProSerHisArgCy 456
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
922 ATTTTAAATAATCGTAGAGTACTGGGAAAGAGAGAGTGTGCTACAT 971
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
456 sAspValPheLeuLysProArgGluArgGluLysLeuArgGluLeuPheP 473
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
972 AGCTCTCCTACCTGCCCAAGAAAGGACCTTTTACGTGAGAGTTGCC 1021
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
473 roGluLeuSerIleGlnPheSerAspSerValArgSerHisProPhe 489
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1022 CAGAGTTAAATTTGGTATTTGTTGATTCGGTGCATAATGAGCACCAATG 1071
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
490 AlaAsnAlaMetArgSerCysPheAsnGlyIlePheSerArgArgCysGl 506
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1072 TGCATAGCGTACGCTATTTGTTTAACTGCTATTGTCATCAGCGTATAA 1121
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
506 yAsnValCysPhePheAspIleGlyGlySerPheThrTyrHisValLys 523
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1122 GAATGTCAAATTTTGGATGTCGGTGGCAGCGAGAGTTATCAGCTGAACA 1171
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
523 laGlyHisValAsnCysHisValCysAsnProValLeuAspValLysAsp 539
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1172 TGGGGAACGTTAACGCTCATGTTGTTAACCCCATGGTTGATTGGAAGGAC 1221
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
540 ValLysArgArgIleAsnGluIleLeuPheLeuSerThrAlaGlyClys 556
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1222 GCGGACGACGGGTACGAGAGTCTATGTACCTCTCAGAGCGCGCTACAA 1271
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
556 pSerTyrValSerAspLeuThrGluAlaAlaSerLysSerValS 573
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1272 GGTGCTAACGTGGAGTCGTTGTAAGGAGGTTGCCGCTGCTAAAGTAA 1321
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
573 erTyrCysSerArgGluSerGlnAsnCysAspSerArgAlaAspAlaGly 589
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1322 CATCTCGCGGCGACGTTTCGAGGATTTGGTGTGAGCGCTGATCAGGG 1371
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
590 PheMetValAspValTyrAspIleSerProGlnGlnValAlaGluAlaMe 606
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1372 TTTTGTGTGATGATACGATATGTCCTGCGACGAGTTGGCCGATGGAAT 1421
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
606 tAspLysLysGlyAlaLeuValPheAspIleAlaLeuMetPheProValG 623
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1422 GGAGAAGAAGATATGCCGACTTTTACGTGCTGCTTAATGTACCCCGTAG 1471
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
623 luLeuLeuTyrGlyAsnGlyGluValTyrLeuGluGluLeuAspThrLeu 639
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1472 AGCTGTTGCGAGCTGACGGTGAATTTTACCTTGAACCGCTAGACGTGTTA 1521
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
640 ValLysArgGluGlyAspTyrLeuAlaTyrAsnValGlyGlnCysGlyCl 656
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1522 GTGAAGAGAGAGGGCGATTTGTGTGCGCTATACGTGGGCGTTACAGGCGA 1571
```

656 uMetTyrGluHisSerPheSerAsnValSerGlyPhePheThrPheSerT 673
 1572 GAGCTACACTCATAGTATTCGTTGATATCATCCGCTTATTTCACCGCTCTCG 1621
 673 yValArgThrSerSerGlyAsnValPheLysLeuGluTyrGluGlyTyr 689
 1622 CTGTGGTTCGTAGGGGTGGTACTGTTTCAAAGTGCAAATATGAGGGTTAC 1671
 690 ArgCysGlyTyrHisHisLeuThrMetCysArgAlaGlnLysSerProG1 706
 1672 AGATTGGGGTACCATCATATATCTTGTGCTTAGCGACAGCTCTAGTGC 1721
 706 yThrGluValThrTyrArgSerLeuValProSerPheValGlyLysSerL 723
 1722 TAGCGTTAGTGTGCAAGAGGCACATCCGTACCTCTCCACGGAAAGACGA 1771
 723 euValPheIleProValValAlaGlySerSerValSerPheLysThrIle 739
 1772 TAGTTATGATTCGGAGATTGAATATCGCGTGATATCCTTCGAGCGTATG 1821
 740 ValLeuAspSerAspPheValAspArgIleTyrSerTyrAlaLeuAsnTh 756
 1822 GGTGACGATACGTATTCGTCGATAGGATTATCGTAGCTAGCTCTAAACAC 1871
 756 rIleGlyThrPheGluAsnArgThrPheGluTyrAlaValAlaGlyAlaVal 773
 1872 GATCGCCAACTCGAGAAATCGTAGCTTCGAGTAGCGCGTAGGCAACGTTCT 1921
 773 rgSerGlnLysThrHisValIleThrGlySerArgValValHisSerLys 789
 1922 GGGCACAAAGACCGATGCTATTACTGGACGCGTGGTGATAGTAGAGA 1971
 790 ValAspIleSerProAspAspMetTrpGlyLeuValValAlaValMetAl 806
 1972 GTAGATATAACGAGTAATGACTTGTGGGTTGTGTTGCGGTATGACAC 2021
 806 acGlnAlaIleLysAspArgAlaLysSerIleArgSerTyrAsnPheIleL 823
 2022 TCAAGCCGTGAAGATAGCGCAAGAGCGCTTGATACGTATTTCTTGTTAA 2071
 823 ysAlaSerGluGlySerLeuAlaGlyValPheLysLeuPhePheGlnThr 839
 2072 AGGCTTGGGTGGAACACTTTTGTGCACTAATAAATGTTATG..... 2112
 840 ValGlyAspCysPheSerAsnAlaValSerValTyrAlaLysAlaMetVa 856
 2113 GTGACACACTGTTGAGGTACAGTACCACGCTCTTAGAGGCTTATGTCG 2162
 856 hIisAsp.....AsnPheAsnValLeuGluThrLeuMetSerMetp 870
 2163 GAGTTCCTCGGCGTTAAATTTTGTACTTTTCAGCTCTCTGTTGGCAGTGC 2212
 870 roArgAlaPheIleArgLysValProGlySerValValThrIleCys 886
 2213 CTGAAAAATTACATATATAGGTTACCGAAGACGCTGTTCTTTGAATTTACGA 2262
 887 ThrSerGlyAlaSerAspArgLeuLeuArgGlyAlaPheAspIleSe 903
 2263 ACGCTATTACCGTTAGTGACCGCTGAATTTTGACTGTCGACCTCGGGAAC 2312
 903 rLysGluThrPheGlyArgLysLeuLysAsnSerArgLeuArgValPheS 920
 2313 ACGTCAACGACTACTTTCAGGATATCCGTAGCAAAAGCTATAAGTACGTTT 2362
 920 erArgAlaIleValGluAspSerIleLysValMetLysAlaMetLysThr 936
 2363 CGAAGGTTCTCGTTAAAGATGCAATAAGCCCTAATGAAGAAAAAGCGGTGT 2412
 937 GluAspGlyLysProLeuProIleThrGluAspSerValTyrAlaPheI1 953
 2413 TCCGATGTGTAGGATCGCGATATTAACCTGAAGAGACGCTCGTGCATTTAT 2462
 953 eMetGlyAsnValSerAsnValHisCysThrArgAlaGlyLeuLeuGluGlv 970

[illegible]

3291 GGCTGTATACATATGCGGGGACCACCAAGGTAGCTATGGCTCTGG 3340
1189 lyThrLeuLeuGlyAlaArgAlaSerAlaAsnThrTrpLysPheGlyGly 1205
:::|||||
3341 CGACGTTGTTGGGTTAAATAAAGCGTAGTCCATGG..... 3378
1206 PheSerAsnAsnTrpCysAlaValProGluValValTrpArgGlyLysSe 1222
:::|||||
3379ACCGTACGTAATGTCGTCGTTAATGGT..... 3405
1222 rValSerSerLeuLeuProIleThrLeuGlyValSerLeuIleAla 1239
||| |||||
3406 .GTAGTTTCATTACTG.....GGACTATACT 3430
1239 rgGlyLeuLeuAsnAspThrIleProGlnLeuAlaTyrValProVal 1255
||| ||| ::||| ||| ::| |||
3431 ATCGTACACTGCCCGAAACGAGCCCGGTATTATAATAGTTCGCCCTGTG 3480
1256 GluGlyArgAsnValTyrAspGluThrLeuArgTyrTyrArgAspPheAs 1272
::: |||||
3481 AATACTTGTAAATGTTTACGAGCGGTGCGAGGCTTACTTT..... 3519
1272 pTyrAspGluGlyAlaGlyProSerGlyThrGlnHisGluAlaValProG 1289
|||||
3520GACCCGTCC..... 3528
1289 lyAspAspAsnAspGlySerThrSerSerValSerSerTyrAspValVal 1305
3528 3528
1306 ThrAsnValArgAspValGlyIleSerThrAsnGlyGluValThrGlyG 1322
|||||
3529AGGCAC.....TCACTAACGCGGAAA 3548
1322 uGluGluThrHisSerProArgSerValClnTyrThrTyrValGluGluG 1339
:::|||||
3549 CGACGATCGAACTCT.....GATGATG 3571
1339 luValAlaProSerAlaAlaValAlaGluArgGlnGlyAspProSerGly 1355
:::|||||
3572 AGCCTAGTTCGCTAGTGTGTGCA..... 3597
1356 SerGlyThrAlaAspAlaMetAlaPheValGluSerValLysGlyVa 1372
:::|||||
3598CTCGATTGCGTTAAA..... 3612
1372 lAspAspValPheHisGlnGlnSerSerGlyGluThrAlaArgGluValG 1389
3612 3612
1389 luValAspGlyLysGlyLeuLeuProGluSerValValGlyGluAlaPro 1405
3612 3612
1406 ThrGlnGluArgGlyArgAlaAlaAspGlyAsnThrAlaGlnThrAlaVa 1422
::: ||| ::||| ||| ::
3613AGCCTGATGATTCACAGATACACAGTATAAT 3644
1422 l...AsnGluGlyAspArgGluProValClnSerSerLeuValSerSerp 1438
::: |||
3645 GTATAGGACAAAGAAATCCACCCCTACAGTCAGGTATC.....C 3685
1438 roGlnAlaAspIleProLysValThrGlnSerGluValHisAlaGlnLys 1454
|| ::|||
3686 CGACTTGCTCATATCGCCTGTGTAACGAACACGCGGTTTCATACTGGCCCG 3735
1455 GluValLysGlnGluValProLeuAlaThrValSerGlyAlaThrProIl 1471
|||
3736 GAG.....AT 3740
1471 eValAspGluLysProAlaProSerValThrThrArgGlyValLysIleI 1488
|||||
3741 AGTGAGTTGGACCTTCGCCCA.....C 3763

1488 leAspLysGlyLysAlaValAlaHisValAlaGluLysLysGlnValGln 1504
:::|||||
3764 TTTGATAAAGAAAGCGTATTGAACGCGTA..... 3792
1505 ValGluGlnProLysGlnArgSerLeuThrIleAsnGluGlyLysAlaGl 1521
3792 3792
1521 yLysGlnLeuCysMetPheArgThrCysSerCysGlyValGlnLeuAspV 1538
3792 3792
1538 alTyrAsnGluAlaThrIleAlaThrArgPheSerAsnAlaPheThrPhe 1554
3792 3792
1555 ValAspAsnLeuLysGlyArgSerAlaValPhePheSerLysLeuGlyG 1571
3792 3792
1571 uGlyTyrThrTyrAsnGlySerHisValSerSerGlyTyrProArgA 1588
:::|||||
3793AATACGCACTCAAGCTCG..... 3810
1588 laLeuGluAspIleLeuThrAlaIleLysTyrProSerValPheAspHis 1604
:::|||||
3811GAGATAGTGGAGTTGGAACCTCTACCACCTGTT..... 3843
1605 CysLeuValGlnLysTyrLysMetGlyGlyValProPheHisAlaAs 1621
3843 3843
1621 pAspGluGluCystTyrProSerAspAsnProIleLeuThrValAsnLeuV 1638
3843 3843
1638 alGlyLysAlaAsnPheSerThrLysCysArgLysGlyGlyLysValMet 1654
|||||
3844AAGGAAACCGGTTGAA 3861
1655 ValIleAsnValAlaSerGlyAspTyrPheLeuMetProCysGlyPheG 1671
:::|||||
3862 TCGGTGAAT.....ATGCCG..... 3876
1671 nArgThrHisLeuHisSerValAsnSerIleAspGluGlyArgIleSerL 1688
|||||
3877CACATACCCATTAGTGGCGAGCTCGATACAAACCGGCTCTA 3919
1688 euThrPheArgAlaThrArgValPheGlyValGlyArgMetLeuGln 1704
::: |||
3920 GCAGCTGTACTCGCCACCGCAGAGTTT..... 3948
1705 LeuAlaGlyGlyValSerAspGluLysSerProGlyValProAsnGlnG 1721
:::|||||
3949ATACCTACAAAACC 3962
1721 nProGlnSerGlnGlyAlaThrArgThrIleThrProLysSerGlyGlyL 1738
:::|||||
3963 GCGTACAATCCGCGATTATCTGATTCCGTTGTCCT.....TCGA 4003
1738 ysAlaLeuSerGluGlySerGlyArgGluValLysGlyArgSerThrTyr 1754
||:|||||
4004 AGTCAATGAGT.....ACTACTTTG 4023
1755 SerIleTrpCysGluGlnAspTyrValArgLysCysGluTrpLeuArgAl 1771
|||||
4024 AGTATGATGAGGACAGCTCGCGGTCTCGCGCAACAAATGCTATTGGG 4073
1771 aAspAsnProValMetAlaLeuGluProAspTyrThrProMetThrPheG 1788
:::|||||
4074 ACGAGACCTGTCTTCTTGTATGAGCAGAGTGAGCCTGTCTATTGTAG 4123

1788 luValVallysThrGlyThrSerGluAspAlaValValGluTyrLeuLys 1804
 :
 4124 GAGACCTAGGTGCATTTGGTCGAAGCGCCGGTGCTGAATCATCTTG 4173
 :
 1805 TyrLeuAlaIleGlyIleGluArgThrTyrArgAlaLeuLeuMetAlaAar 1821
 :
 4174 TACGTGCTACTGGTTATTCAAGCTGTACAGATCCTTAGTGAACGCCGC 4223
 :
 1821 GasnIleAlaValThrAlaGluGlyValLeuLysValProAsnGlnV 1838
 :
 4224 GAGACTGGAATCGTTAAATGATCAAAGCGCGTCTGTTTCCCGCAGTGA 4273
 :
 1838 alTyGluSerLeuProGlyPheHisValTyrLysSerGlyThrAspLeu 1854
 :
 4274 TACATGACGAGTTGCCAGGTATATCGGAGTA...AGCCCCAACGCATATT 4320
 :
 1855 IlePheHisSerThrGln...AspGlyLeuArgValArgAspLeuProTy 1870
 :
 4321 ATNTTTTTAAGTCGACGGAGAGACGTGGAGATTGACTCACCTGCGGTA 4370
 :
 1870 rValLeuIleAlaGluLysGlyIlePheThrLysGlyLysAspValAspA 1887
 :
 4371 TGTTTTCTCTGFAAATAGAAAGAGATTGTGACACCGCTTCGCGCTCGAAGA 4420
 :
 1887 laValValAlaLeuGlyAspAsnLeuPheValCysAspAlleLeuVal 1903
 :
 4421 TGCCCGATCCCGCTACCGACGAGATATCTGGTGTGTGATGAGATGGTTCCA 4470
 :
 1904 PheHisAspAlaIleAsnLeuIleGlyAlaLeuLysValAlaArgCysGI 1920
 :
 4471 TTCCATACGCACTGAATTTGGTTCGCGCCATTCAATTTATGCGCTGCGCG 4520
 :
 1920 yMetValGlyGluSerPheLysSerPheGluTyrLysCystyrAsnAlap 1937
 :
 4521 AAAGCTAGCGAAAATTTTTCAAAAGTAACCAAGTATGCGTTAATGCCCC 4570
 :
 1937 roProGlyGlyLysThrThrLeuValAspGluPheValLysSer 1953
 :
 4571 CGCCTGGAGGAGAAAGACAACAAATTTGATTCAGGAGTACTTTAAGGAT 4620
 :
 1954 ProAsnSerThrAlaThrIleThrAlaAsnValGlySerSerGluAspII 1970
 :
 4621 AGATCTGGCAGTTGATTATGACC GCCCANTCGTTGATGCGCTATAGATAT 4670
 :
 1970 eAsnMetAlaValLysLysArgAspPro..... 1979
 :
 4671 CAATCATACCATTTAGAGCTATAGACGCTGTCTAATGCGAGCAACGCCGCTT 4720
 :
 1980AsnLeuGluGlyLeuAsnSer.....AlaThr 1988
 :
 4721 CGAAACAAGCTCAGCGCGCTGGAATCGATAGATAATTAATTTGTGCTCGT 4770
 :
 1989 ThrValAsnSerArgValValAsnPheIleValArgGlyMet.....Ty 2003
 :
 4771 ACAGTTAACTCACAATTTATGAAC.....TGCAAAAGGAGTAAATGAATTA 4814
 :
 2003 rLyArgValLeuValAspGluValHisMetMetHisGluLeuLeuG 2020
 :
 4815 TACCTCGCGCCCTAGTTGACGAANAATGATTATGATGCCAAGAGGCTGTGTA 4864
 :
 2020 InLeuGlyValPheAlaThrGlyAlaSerGluGlyLeuPhePheGlyAsp 2036
 :
 4865 TCTTGGGGGTATTTCAGTTCTCGAGCAAGAACAGCATATATTTCTACGGAGAT 4914
 :
 2037 IleAsnGlnIleProPheIleAsnArgGluLysValPheArgMetAspCy 2053
 :
 4915 APTCAACCAAGATCCCCTTCATAAATAGAGAGAAAGTGTGTTTTACTCTAAGGA 4964
 :
 2053 saValPheValProLysLysGluSerValValTyrThrSerLysSert 2070
 :
 4965 GGGTGTGTACTGTCACGGTAAAGATGAAATTTATTTACATACAGAGTCCT 5014
 :
 2070 yrArgCysProLeuAspValCysTyrLeuLeuSerSerMetThrValArg 2086

5015	ACAGATGTCCTGCCCATGTTTGTATGTGGTCAAGCTCAGTCAAGCGCAA	5064		
2087	GlyThrGluLysCysTyProGluLysValValSerGlyLysAspLysPr	2103		
5065	GCTGGGTCTAATCGCTACTGAAGGGTGTCATGCACACCAGCGTAGT	5114		
2103	oValValArgSerLeuSerLysArgProIleGlyThrThrAspaspVala	2120		
5115	GGTGTACCGCAGTTTATCCAAACGCCGGTAGTGTACCGGAACAAGTGA	5164		
2120	IaGluIleAsnAlaaspValTyrLeuCysMetThrGlnLeuGluLysSer	2136		
5165	TACAATTGGAAGCTCACGCTTATATAAACATTCAAGCAGGAGTGTAAAGA	5214		
2137	AspMetLysArgSerLeuLysGlyLysGluThrProValMetTh	2153		
5215	AAAGTCGTGAGGGCGCTACGAGCTGTAGCAGAGGGGATAAGCTGTTTAC	5264		
2153	rValHisGluLaGInGlyLysThrPheSerAspValValLeuPheArqT	2170		
5265	AAGCATGAGCGCAGGTATCATCTTTTGGCGGGTCGTGTATGTAGAT	5314		
2170	hrLysLysAlaAspAspSerLeuPheThrLysGlnProHISileLeuVal	2186		
5315	TAAGTGCCACTGACCATCCGTTTTTCTCTGAGCCCTCACATTTTACTT	5364		
2187	GlyLysSerArgHisThrArgSerLeuValTyrAlaAlaLeuSerSerLy	2203		
5365	GCACCTCTCCAGACATACACATCCCTGTCTATGCCACTCTTAGTAGTAA	5414		
2203	sLeuAspLysValGlyThrTyrIleSerAspAlaSerProGlnSerV	2220		
5415	CTTACCCACACAGGTAGTGGCGGCATAGACTCAGTTACGCGTAGGAGG	5464		
2220	alSerAspAlaLeuLeuHisThrPheAlaProAlagLycyspheArg	2235		
5465	TAAGTGATACGCTACTTAAGACCTTTGTGGCGTCGGCGTTATTTCGA	5511		

```
seq_name: gb_vil.AF195822

seq_documentation_block:
LOCUS      AF195822       12394 bp        RNA                VRL          09-MAR-2000
DEFINITION Grapevine leafroll-associated virus 1 helicase (HEL) and
              RNA-dependent RNA polymerase (POL) genes, partial cds; and p7,
              HSP70-like protein, p55, coat protein (CP), p55 (CPol), p50 (CPd2),
              p22, and p24 genes, complete cds.
ACCESSION   AF195822
VERSION     AF195822
KEYWORDS    .
SOURCE      .
ORGANISM    grapevine leafroll-associated virus 1.
            grapevine leafroll-associated virus 1.
            Viruses; ssRNA positive-strand viruses, no DNA stage;
            Closteroviridae; Closterovirus.
REFERENCE   1 (bases 1 to 12394)
AUTHORS     Fazeli,C.F. and Rezaian,M.A.
TITLE       Nucleotide sequence and organization of ten open reading frames in
            the genome of grapevine leafroll-associated virus 1 and
            identification of three subgenomic RNAs
JOURNAL     J. Gen. Virol. 81 Pt 3, 605-615 (2000)
MEDLINE     2041373
REFERENCE   2 (bases 1 to 12394)
AUTHORS     Fazeli,C.F. and Rezaian,M.A.
TITLE       Direct Submission
JOURNAL     Submitted (18-OCT-1999) Plant Industry, CSIRO, Hartley Grove,
            Urrbrae, Adelaide, SA 5064, Australia
FEATURES    Location/Qualifiers
             source          1..12394
                        /organism="grapevine leafroll-associated virus 1"
                        /db_xref="taxon:47985"
             gene           <1..1199
             CDS             <1..1199
```


alignment_block:

US-09-301-906-15 x AF195822

Align seg 1/1 to: AF195822 from: 1 to: 12394

```
1847 ValTyrLysSerGlyThrAspLeuIlePheHisSerThrGlnAspGlyLe 1863
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
9  GTGACGATGCCAATTTCAAGGTGGTGAATGCACCGGAGGACGGAAC 58

1863 uArgValArgAspProTyrValLeuIleAlaGluLysGlyIlePhe 1880
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
59  TACTATTTTCGACCTGGAATACCTCTTCTCGTATCTACAGGCACCTTTG 108

1880 hrlYsGlyLysAspValAspAlaValAlaLeuGlyAspLeuPhe 1996
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
109  TGGCAGCGGGGAATTTGCAAGCGCTTATCACGACAAGATCGGTACTG 158

1897 ValCysAspIleLeuValPheHisAspAlaIleAsnLeuIleGlyAl 1913
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
159  GTGTGCGATGAAGTCTGCTTTTCATGATGCANTGAATTTGCGCGGATG 208

1913 aLeuLysValAlaArgCysGlyMetValGlyGluSerPheLysSerPhe 1930
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
209  CGTTAGTTGGCGAAAGACATGTCGCGAGAATATATCAAGGATGTAA 258

1930 luTyrLysCysTyrAsnAlaProProGlyGlyLysThrThrLeu 1946
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
259  GAATTTCTCGGTGAATCTACCGCTGTGTGAGGTAACACACGACGCTG 308

1947 ValAspGluPheValLysSerProAsnSerThrAlaThrIleThrAla 1963
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
309  GTTGATGAATACTTTGGAGGAAGACGAGCTAAAATCGCAGCAGCCAA 358

1963 nValGlySerSerGluAspIleAsnMetAlaValLysLysArgAspPro 1980
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
359  CACCGGAGGTAGCCGACATCAATCGGCTATTCAGCAAGAGAGGAA 408

1980 snLeuGluGlyLeuAsnSerAlaThrThrValAsnSerArgValVal 1996
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
409  AGAAGAGCCAGATTTAGTAGCAAGACCGGAAATTCATGGGTCAATAAC 458

1997 PheIleValArgGlyMetTyrLysArgValLeuValAspGluValHis 2013
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
459  TGCATCCACGTCGCACTCGCAGCTTGACCTTATCGATGAGGTTTATAT 508

2013 tMetHisGlnGlyLeuLeuGlnLeuGlyValPheAlaThrGlyAlaSer 2030
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
509  GCTACATAAGGGAATTTTCAGCTTACTGTGCTCTATGGGAGTGAAG 558

2030 luGlyLeuPhePheGlyAspIleAsnGlnIleProPheIleAsnArg 2046
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
559  AGGTGATAGGTACGGGACAGCAACCAATTCGTTTCAACACAGAG 608

2047 LysValPheArgMetAspCysAlaValPheValProLysLysGluSer 2063
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
609  AGACATCTGCTGACTCCAATGAAGCTGGAGTTCGCTCAGGAGCAGAT 658

2063 lValTyrThrSerLysSerTyrArgCysProLeuAspValCysTyrLe 2080
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
659  AGATTATACTGACATTTCTTACCGATGTCACGAGATGTTGTATGTTT 708

2080 euSerSerMetThrValArgGlyThrGluLysCysTyrProGluLysVal 2096
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
709  TGTCTTCGATGACTGATGCAGCTGGGAAGAAATGTACCCAAACGAGTG 758

2097 ValSerGlyLysAspLysProValValArgSerLeuSerLysArgPro 2113
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
759  TTCCCGGAATGGAGACGTGAGACTCTACGCTCGTTCCGAGAAGGTGCC 808

2113 eGlyThrThrAspValAlaGluIleAsnAlaAspValTyrLeuCys 2130
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
809  CGCAACCCCGGAGACGCTTTACTTATGACGCGGACGTGTATCTGACGA 858
```

```
2130 etThrGlnLeuGluLysSerAspMetLysArgSerLeuLysGly..... 2144
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
859  TGACTCAGACGAGAAAGGCTGAATGCAGAGGCTGTGGCAAAATGGAA 908

2145 .....LysGlyLysGluThrPro...ValMetThrValHisGluAlaGl 2158
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
909  GTTGTGGCTGGGAAGAAAGACCGGATGTGATTAATCTCATGAGCGCA 958

2158 nGlyLysThrPheSerAspValValLeuPheArgThrLysLysAlaAsp 2175
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
959  AGCTAAGACATACGAAACGCTGCTTGTGTGAGACTGAAGAGCGCGATG 1008

2175 spSerLeuPheThrLysGlnProHisIleLeuValGlyLeuSerArgHis 2191
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1009  ATCCAATTTCTCCGAAACACACATGTTGTCGCGCTGTGAGGCAC 1058

2192 ThrArgSerLeuValTyrAlaAlaLeuSerSerLysLeuAspLysVal 2208
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1059  ACCGCGATGATGAGTACGCGTGTGAGTTCGAAGATGACTGATACGAT 1108

2208 lGlyThrTyrIleSerAspAlaSerProGlnSerValSerAspAlaLeu 2225
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1109  ATCGAAGCTAATTCATGCTGACGAGTCCCGTAAGTAAGTACGCTTTGT 1158

2225 euHisThrPheAlaProAlaGlyCysPheArg 2235
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1159  TGCAGCAGTTGCAGCGGACGATGCTTTCGA 1190

seq_name: gb_v1:AF039204

seq_documentation_block:
LOCUS AF039204 15000 bp RNA VRL 22-MAY-1998
DEFINITION Grapevine leafroll-associated virus 2 methyltransferase/helicase
polyprotein gene, partial cds; and RNA-dependent RNA polymerase,
putative transmembrane small hydrophobic protein, 65 kDa chaperone
protein, 63 kDa protein, 25 kDa diverged coat protein, 22 kDa coat
protein, 19 kDa protein, and 24 kDa protein genes, complete cds.
ACCESSION AF039204
VERSION AF039204
KEYWORDS AF039204.1 GI:3123909
SOURCE grapevine leafroll-associated virus 2.
ORGANISM grapevine leafroll-associated virus 2
VIRUSES; ssRNA positive-strand viruses, no DNA stage;
Closteroviridae; Closterovirus.
REFERENCE 1 (bases 1 to 15000)
AUTHORS Zhu,H.Y., Ling,K.S., Goszczynski,D.E., McFerson,J.R. and
Gonsalves,D.
TITLE Nucleotide sequence and genome organization of grapevine
leafroll-associated virus-2 are similar to beet yellows virus, the
closterovirus type member
J. Gen. Virol. 79 (Pt 5), 1289-1298 (1998)
89264507
2 (bases 1 to 15000)
Zhu,H.Y., Ling,K.S. and Gonsalves,D.
Direct Submission
Submitted (18-DEC-1997) Plant Pathology, Cornell University, New
York State Agricultural Experiment Station, Geneva, NY 14456, USA
FEATURES
source
1..15000
/organism="grapevine leafroll-associated virus 2"
/db_xref="taxon:64003"
<1..7423
/function="replication and unknown functions"
/note="larger than 277 kDa; ORF1a; contains domains for
two papain-like leader proteases, a methyltransferase and
a helicase; identified by sequence comparison"
/codon_start=2
/product="methyltransferase/helicase polyprotein"
/protein_id="AAC40855.1"
/db_xref="GI:3123910"
/translation="ADYVAMLRVYCGGKFPFLVMSRVTPDRCYLAHMYLCAPYCR
PFRESYDALGMWPTVARLRACVEKNFGVACGIALRGVYTSRNVYHCDYDSAYVKYFR
```


BASE COUNT	3877 a.	3050 c	3820 g	4253 t	/note="similar to 3'UTRs of other monopartite closteroviruses"
ORIGIN					

```
alignment_scores:
  Quality: 794.50
  Ratio: 0.683
  Percent Similarity: 46.709
  Length: 2492
  Gaps: 113
  Percent Identity: 21.589
```

```
alignment_block:
US-09-301-906-15 x AF039204
```

Align seg 1/1 to: AF039204 from: 1 to: 15000

152 GlnAspGluInLysProLysArgGlnAlaSerHisTrpAlaValLysPr 168
::: ||| |||||::: :::: |::|::|::|::|::|::|::|::|::|::|::|::|
473 CAATCGACGGTTAAGCAACAACGTGAAGATGCGCACCCAAACCGTAGAGGA 522

168 oThrAlaValGlyValHisValProLeuProLysLysGlnGluAlaLeuG 185
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
523 ACGGCGCCGCCGCGCATCGTCGAGCAACCAGCAGAAAAGAGGATAGATG 572

185 LuPro...AlaGlnSerValProGlnGlnSerLeu..... 195
||::: |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
573 AGAAAGGTTGCGGCAGAGTCTCTAGTGAGGTTTTTCGCATCTCCTGGTC 622

196GluGluLysAlaAlaLeuThrPheGlyLeuPhePheSe 208
|||::: |||::: |||::: |||::|::|::|::|::|::|::|::|::|::|
623 GCACCACTTAAGCAAGTTTAGGAGGAAGTAGCTGCGGACTCTACGGTT 672

208 rLysGlyGlyGlyAspGluSerAspAlaValIleLeuArgLysGlyLysL 225
::: |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
673 TCGCGTTGCGGCTGATATGGAT..... 694

225 euPheAsnArgAlaLeuAsnValProIleAspValLysAsnThrPheVal 241
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
695 ...TTTCATCGCTCGTTCTCCACCCAAGCGGC...CACCGCTTGCTGGTG 739

242 TrpAlaLys.....IleTrpAspGluAlaSe 250
|||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
740 TGCCCCCGCTCGACCGCGAGCGTGCCCTTGAACTTTACTCACCATTCAA 789

250 tArgArgArgGlyTyrrPheTyrvallYs.....AspArgAlaValL 264
::: ||| |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
790 AAACCTTTTTCGTTACGATGCTTGCCCTGTTCTGAGACTATGACGAGA 839

264 ySPhePheProIleValArgAlaThrIleGluAspPheIleVal 280
|||::: |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
840 TGTTTTCTTCGCGCGCGCGCGCTTCCCTTA..... 874

281 AsnThrAlaProGlyCysAspValAlaLeuProArgIleGluLeuTrpse 297
|||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
875GTTTGTGACTAGAAITAGATAc..... 898

297 rMetargGluargAlaPheValCysThrThrySLySglyTrpCysTrpPheA 314
::: |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
899CCGAACGGGTTTTGTTACTTGG 920

314 snAsnGluArgLeuArgGlyLuIleTyraArgArqCysPheSerSer 330
::: ||| |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
921 CTCACTCGCGGTACCGTGC GGTTTCTCTTAAGGGGTTTTGTATCCGAG 970

331 SerPheSerIleGly.....Ph 336
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
971 CGTTTCGACATCGGTGCTTTCCCACCGCGCAAGCTCAGAAACCGGAT 1020

336 eLeuMethIstLeuGlyPheArgSerLeuLysValIleArgPheAlaGlyt 353
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1021 GTTTCGGAGCTTGTGTAAGAAAGTTTA..... 1048

353 hrASnIleLeuHisMetProSerLeuAsnGluArgThrPheGlyTrp 369

[illegible]

[illegible]

925 GluAspSerIleLysValMetLysAlaMetLysThrGlu.....As 938
 ||||| : : : ||| : : :
 2708 GGAGATTTCGATTCCGGCTAAACTCCGACGAGAGATAGTGGTCGATATTGA 2757
 938 pGlyLysProLeuProIleThrGluAspSerValTyrAlaPheIleMetG 955
 ||||| : : : ||||| : : :
 2758 CTTCTAAGCCCTTACCACGACAG.....GTGG 2783
 955 LysAsnValSerAsnValHisCys.....ThrArgAlaGly 966
 ||||| ||||| : : : : : ||||| |||||
 2784 GTAATTCTGTCATCCGAAACCGCGATACCGGGAAGTTCACAGACCGCCGT 2833
 967 LeuLeuGlyGlySerLysAlaThrValValSerValSerLysGlyLe 983
 ||| ||||| : : : ||||| : : :
 2834 TTTCGGTGGTGGTTCTAGAACGGGCTTGTGGGAG..... 2869
 983 uValAlaArgGlyAlaAlaThrLysAlaPheSerGlyIleThrSerPhe 1000
 ||||| : : : : : ||||| : : :
 2870TTCTTTCACCTTCG 2882
 1000 heSerThrGlySerLeuPheTyrAspArgGlyLeuThrGluAspGluArg 1016
 : : : : : ||||| : : : : : ||||| : : : : :
 2883 TCTGTGATTCTGCCTTGGCTCTTTTCAAAATACGGCAGCGATCAACAACGG 2932
 1017 LeuAspAlaLeuValArgThrGluAsnAlaIleAsnSerProValGlyI 1033
 : : : : : ||||| : : : : : ||||| : : : : :
 2933 ATCAAGTCTTACGTGGT.....TTCTTGGACTCGCGGCTCTCATTT 2973
 1033 eLeuGluThrSerArgValAlaValSerLysValValAlaGlyThrLysG 1050
 : : : : : ||||| : : : : : ||||| : : : : :
 2974 CTTGGATTACAACACTACGATAATCTATCTGTTTATCTCGCAGTGCCTTCG 3023
 1050 LuPheTrpSer..... 1053
 ||||| : : : : :
 3024 AAGGTTATTTCGGTGTATGTTCCGGTTTTTGGCAATCCGGCGAGCTTATCT 3073
 1054GluValSerLeuAsnAs 1059
 ||||| : : : : :
 3074 AGTCGTGTCGTAGCGCGGTGTGCTGTGAAAGAAGTTGCTACTCTCATG 3123
 1059 pPheThrPheValLeuArgAsnLysValLeuIleGlyIlePheValA 1076
 : : : : : ||||| : : : ||||| : : : |||||
 3124 CCGCAACGCGACGGTTCTTAAAGCCAAGTGTATGATT..... 3160
 1076 IaSerLeuGlyAlaAlaProIleAlaTrpLysTyrArgArg...GlyIle 1091
 : : : : : ||||| : : : ||||| : : : ||||| : : :
 3161 ..ACCTTCGACGCGCGGTGTGCTGATGATGTTTAAATAGTCGCGGTTT 3208
 1092 AlaAlaAsnAlaArgArgTyrAlaGly..... 1100
 : : : : : ||||| : : : : :
 3209 TCAGGCGAGGTTCGGAGATATAATCGTATATACATCGTTACACGCAAGT 3258
 1101SerSerTyrGluThrLeuSerS 1108
 ||||| : : : : :
 3259 ATTGTTTGACACTATCTTTTGTAGGACAGCAGTTACTACCCATAGAAG 3308
 1108 eLeuSerSerGlnAlaGlyGlyLeuArgGlyLeuThrSerSer... 1123
 ||||| : : : : : ||||| : : : ||||| : : :
 3309 TTTCTGAGTTCGGCGATATCCGGTGCTATPCGTACACACTTTTCTCCTCGGGC 3358
 1124ThrValSerGlyClySerLeuValValArgArgGlyPheSe 1137
 : : : : : ||||| : : : : : ||||| : : : ||||| : : :
 3359 TCGTCCATAAGTTTAAACGCCTTCTTACTTCAAAATACCAAGAGATTCTC 3408
 1137 rSerAlaValThrValThrArgAlaThrValAlaLysArgGlnValPro 1154
 ||||| : : : : : ||||| : : : ||||| : : :
 3409 CCTAGAGGTTGTCGTC.....CGGAAGTCTGTGTC 3437
 1154 euAlaLeuLeuSerPheSerThrSerTyrAlaIleSerGlyCysSerMet 1170
 : : : : : ||||| : : : : : ||||| : : : ||||| : : :
 3438 GAGTCACGATCGTTTGTAGACACACA...GCCACCGCAGCGCT...ATA 3481

```

1171 LeuGlyIleThrAlaHisAlaLeuProArgHisLeuMet..... 1183
1184 ..... 1192
3482 CGTGGGGTTTCCCAA...ATGTGTCCTACTTACTTGTGGAAATAC 3528
1184 .....PhePheGlyLeuGlyThrLeuL 1192
3529 GGGTAATGTGGCTTACCAGTCAGCTTTCATTCGCGGGTGGTGCCTCTTT 3578
1192 euGlyAlaArgAlaSerAlaAsnThrTrp.....LysPhe 1203
3579 TAGTTAAAGTGTGTGAGCTTAATCTTCATCTTGCCTGAAGATACATTAT 3628
1204 GlyGlyPheSerAsnAsnTrpCysAlaValProGluValValTrp..... 1218
3629 TCCGGTTTATTAAAGCAC.....GGNATCAGTGAATCTCTTTCCTTAG 3672
1219 .....ArgGlyLysSerValSerSerLeuLeuLeuP 1229
3673 TAGTATTCTCAAGTCTTGAAGGTAAGCTTGTGGACGAGTTG..... 3715
1229 rolleThrLeuGlyValSerLeuIleLeuArgGlyLeu..... 1241
3716 .....AATCGATTATTCAAGGGGTTTTTGATTCCAAC 3748
1242 .....LeuAsnAspThrIlePr 1247
3749 AAGCAGCTGTTTAAAGNAGTACTCAGGAAGCGATTCTGACGAGGTCAT 3798
1247 oGln.....LeuAlaTyrValProProValGlu.....GlyArgA 1259
3799 GCAAGTGCCCTGTCCTGTACTGATGCCCTTAAAGCGCGCGGGAAGAA 3848
1259 snValTyrAspGluThrLeuArgTyrTyrArgAspPheAspTyrAspGlu 1275
3849 TTTATAACAATTTTACTAGTCGA.....CGTACCTTTGTGAAGATGAA 3892
1276 GlyAlaGlyProSerGlyThrGlnHisGluAlaValProGlyAspAspAs 1292
3893 GGTCTCTAGCAGCGCGCATGTGAGAGTATTCTCATCGCAGCAGNAGG 3942
1292 nAspGly...SerThrserSerValSerSerTyrAsp..... 1303
3943 TGAAGTCCGGGCTGAAGAGGGGTTCCAGCTATGGCTTCTCAATTTTAG 3992
1304 .....ValValThrAsnValArgAspValGlyIleSerThr 1315
3993 CGTCTCTTCCACCATATGTGGGAGCTCGCGCTTATTGTAAGGTG 4042
1316 AsnGlyCluValThrGlyGluGluGluThrHisSerProArgSerValG 1332
4043 AAGCATGAGTGTGTTGGGAAA.....CT 4065
1332 nTyrThrTyrValGluGluValAlaProSerAlaAlaValAlaGluA 1349
4066 TTTTGAATTTCTA.....TCGCTCAAGCTTCACGAAT 4097
1349 rGlnGlyAspProSerGlySerGlyThrAlaAspAlaMetAlaPheVal 1365
4098 TCAGGACTCGCGTGTGGAAGAATAGACGACGCTG..... 4135
1366 GluSerValLysLysGlyValAspAsp.....ValPheH 1377
4136 .....GGAGTTTACGATTTTTCGCCACGGCATCGTGA 4170
1377 sGlnGlnSerSerGlyThrAlaArgGluValGluValAspGlyLysG 1394
4171 AACGCTCTCATCATAGAGAGTCCGACCAAAATTGAA.....G 4208
1394 lyLeuLeuProGluSerValValGlyGluAlaProThrGlnGluArgGly 1410
4209 AACTTCTCGCGCAGCAGCTGAAAGGT..... 4234
1411 ArgAlaAlaAspGlyAsnThrAlaGlnThrAlaValAsnGluGlyAsp.. 1426

```

```

4235 .....GACAGGAGTCTTCTGTTGACCGATATGAATTACTTTGAGTT 4275
1427 .ArgGluProValGlnSerSerLeuValSerSerProGlnAlaAspIleP 1443
4276 CTCAGAAGACTTCTTACGCTCTATCGAGGAGCCGCTTTCGCTGGATTGC 4325
1443 roLysValThrGlnSer..... 1448
4326 GAGGAGTAGTCAAGAACATCCGATTTTGGCGATTTTGGAAATACGCCAT 4375
1449 .....GluValHisAlaGlnLysGluValLysGlnGluValProLe 1462
4376 AATTGTTTTCGATTCGCAAGCAAGTGTTCGAAACGACCTTTATTCTT 4425
1462 uAlaThrValSerGlyAlaThrProIleValAspGluLysProAlaProS 1479
4426 TGCCTTCGCGCAACTCTCAAGCGCCCTTATCGAGAAA.....TTTAAGG 4469
1479 erValThrThrArgGlyValLysIleLeuAspLysGlyLysAlaValAla 1495
4470 AGGTTTTTCCCTCGT.....AAGGCCAGCTCGTCGCT 4501
1496 HisValAlaGluLysLysGlnValGlnValGluGlnProLysGlnArgSe 1512
4502 ATCGTGGCGGAGTATACTCAG...AGATTCTCCGAAGTCCCATGCGGTGC 4548
1512 rLeuThrIleAsnGluGlyLysAlaGlyLysGlnLysCys..... 1525
4549 GTTGGGTTTGAATAACGAGTTCGTGGTAAATCTTCGCCGATTTGCTAC 4598
1526 .....MetPheArgThrCysSerCysGlyValGlnLeuAspValTyr 1539
4599 CCGCATTAATGAGCGGAAGGTTTCAGGTTTCGTTCTTAGTAGTGTATT 4648
1540 AsnGlu.....AlaThrIleAlaThrArgPheSe 1549
4649 CGCCACTTAGAGTTTCTCATATATGTGTGTTTCAGCGGAGCGACGTGA 4698
1549 rAsnAlaPheThrPheValAspAsnLeuLysGlyArgSerAlaValPheP 1566
4699 AAAGTTTTTTCCTCGTG...TGTTTAAATCGGGTTAAGTCTCCCTTTCT 4745
1566 heSerLys..... 1568
4746 TCGTGCATCGTAGAGCGAAAGCGTCCGAAGAACCTGTCCTCAGCG 4795
1569 .....LeuGlyGluGlyTyrThr... 1574
4796 CGTCGCTTTTATGACGCTATTAAATTTTCTAAGGCAGAGTATGTCTC 4845
1574 ..... 1574
4846 TCCTTCTAATTTCTTTTCTACCTGTTAGTCTCTGACGTTGATCAGATT 4895
1575 .....TyrAsnGlySerHis..... 1580
4896 CCGCATCTGACGGGTTGAAGGTTGTCGCGCAATGACGCTCTTCCAC 4945
1581 .....ValSerSerGlyTrpProArgAl 1588
4946 CTCTCGCTTCGCCCTTGTAGTGCCTCTATCTGTTAGGGTGG..... 4987
1588 aLeuGluAspIleLeuThrAlaIleLysTyrProSerValPheAspHisC 1605
4988 .....GAAGGTTTAAAGCTACTCTTATCGCACCACTTGTATTATTGT 5033
1605 ys.....LeuValGlnLys.....TyrLysMetGlyGly 1615
5034 GTTTGTCATTGGTGAAGTGTGAACGTCCTTATCAAAAGTTCTTGGGGT 5083
1616 ValProPheHisAlaAspAspGluGluCysTyrProSerAspAsnProIl 1632

```

5084 CTTTCTTTCTTTGTG.....CAACCAAT 5106
 1632 cLeuThrVal..... 1635
 15107 CTTTCTCTTTTGGCGGCGATCTCTACAACGGACAGGTTTGTGGAGT 5156
 1636AsnLeuValGlyLysAlaAsnPheSerThrLysCysArgLys 1649
 5157 ATTCCAGAAACTTGTACAGGCTTGAATTTTCTTAAATGCTTGCCT 5206
 1650 GlyGlyLysValMetValIleAsnValAlaSerGlyAspTyrPheLeuMe 1666
 5207 CGCGCCCTGCCTACTC..... 5224
 1666 tProCysGlyPheGlnArgThrHisLeuHisSerValAsnSerIleAspG 1683
 5225AAAGGTTT.....TTTGAGTCGCTGGCGA 5249
 1683 luGlyArgIleSerLeuThrPheArgAlaThrArgArgValPhe..... 1697
 5250 ACAGCACTGTCTCAAAACCGTTCGAAGACTTCTTCGCTGTTTCGTGAAG 5299
 1698GlyValGlyArgMetLeuGlnLeuAlaGly...GlyVa 1709
 5300 ATGCTCAAACTTCGAAAGGGCGAGGTTGCGTGGCGATGGTAGGGTCT 5349
 1709 lSerAspGluLysSer...ProGlyValProAsnGlnGlnProGlnSerG 1725
 5350 CCATCGCGAGAAAGCGGTACCGTCTATCTTAATCGGGTTCGTGACCG 5399
 1725 InGlyAlaThrArgThrIleThrProLysSerGlyGlyLysAlaLeuSer 1741
 5400 ACGGGGTTCAAGACTTTCGTTAAAGATGCAAGAGTTTCAAGCGTTGCGT 5449
 1742GluGlySe 1744
 5450 ACCGAATGAGAACTTAGAAGATTAGATTCTGCGGTGATCGAANAAC 5499
 1744 rGlyArgGluValLysGlyArgSerThrTyrSerIleTrpCysGluGlnA 1761
 5500 CAATGACGC...AGAAATCGTGACACTAAT.....GACGACG 5534
 1761 sPTyrValArgLys.....Cys 1766
 5535 AATTTACGGCCCTGCTCATGACAGATGCAAGAACTCAACCACTTCTGT 5584
 1767 GluTrpLeuArgAlaAspAsnProValMetAlaLeuGlu..... 1779
 5585 TCG.....AAAGCCAACCTCTGCTGTTTGGCCCTGGAAAGGCGAGTGT 5628
 1780ProAspTyrThrProMetThrPheGluV 1789
 5629 TGTGGAAGACGCTATAAAGTCGGAGAACTTCTAAGACGGTTAATGAGA 5678
 1789 aValLysThrGly.....ThrSerGluAspAlaValVal..... 1800
 5679 TGGTGAAGAAAGGAGTACCAACCGACGGAAGAGTGGCCGCTGCTGTGCG 5728
 1801GluTyrLeuLysTyrLeuAlaIleGlyIleGluArg..... 1812
 5729 GACGATGAAGCGGTGAGAAATCTCTGTTGCTGACGACGAGCAGATTC 5778
 1813ThrTyrArg.....AlaL 1817
 5779 GCCTAAGACAGTCAGGATAAGCAATACCTAAATAGTTAAACTCAAGCT 5828
 1817 euLeuMetAlaArgAsnIleAlaValThrAlaGluGlyValLeuLys 1833
 5829 TCGAATTCGGAAGCGCTATTGTTGTGGACCAACAAGGATACCGGGGT 5878
 1834 ValProAsnGlnValTyrGluSerLeuProGlyPheHisValTyrLysSe 1850
 5879 CTAAACGACCGCGTGGAG.....TTTATATATGCAAGA 5916

1850 rGlyThrAspLeuIlePheHisSerThr.....GlnAspGlyL 1863
 5917 ACTTGCTCTTTTCGAATCCACAGAACTGTGCACCTACTACGATCAAC 5966
 1863 euArgVal.....ArgAspLeu 1868
 5967 TGGCATACTCAACTTCGATCGTTCGTAGCACCATGCACGGAAGATGCT 6016
 1869 ProTyrValIleLeuAlaGluLysGlyIlePheThrLysGlyLysAspVa 1885
 6017 CAGCTTACGTACGGGAAGACGGCTCAAGATAGTCAGGGTAAAGAGT 6066
 1885 lAspAla...ValValAlaLeuGlyAspAsnLeuPheValCysAspAsI 1901
 6067 ACCTTTGCACATTAAAGATTTCACCATCAGGATTTCTGTTTGACGGA 6116
 1901 leLeu.....ValPheHisAsp 1906
 6117 AATTTCTATTAAACAGCGCGGAGCGGAAATGTTTATATACGAC 6166
 1907 AlaIleAsnLeuIleGlyAlaLeuLysValAlaArgCysGlyMetValG 1923
 6167AACCTC.....GCGTTCTTGGCGAGTAATTGTTCTTAGCCGG 6204
 1923 yGluSerPhe.....LysSerPheGluTyrL 1932
 6205 CTACCCCTTTTCAAGGAGCTTCGTCTCAGCAATTCGTGCGTCGATATC 6254
 1932 yScysTyrAsnAlaProGlyGlyGlyLysThrThrThrLeuValAsp 1948
 6255 TCCTCTACGAAGCTCCACCGGAGGTGTAAGACGACGACGCTGATTGAC 6304
 1949 GluPheVal.....LysSerProAsnSerThrAlaThrIleTh 1961
 6305 TCGTCTTGAAGGTCTTCAAGAAAGGTGAGGTTTCCACCATTGATCTTAAC 6354
 1961 rAlaAsnValGlySerSerGluAspIleAsnMetAlaValLysLysArgA 1978
 6355 CGCCACAACAAAGTTCGACGTTGAGATCTTAAGAAAGTGGAGAAGGAAG 6404
 1978 sProAsnLeuGlu.....GlyLeuAsn 1985
 6405 TGTCTAACATTGAATGCCAGAAACGTAAAGACAAAGATCTCCGAAAAAG 6454
 1986 SerAlaThrThrValAsnSerArgValValAsnPheIleValArgGlyMe 2002
 6455 AGCATTACACCATCGACGCTTATTTAATGCAATCAC.....CGTGGTTG 6498
 2002 tTyrLysArgValLeu...ValAspGluValHisMetHisGlnGlyL 2018
 6499 TGATGCAGACGTTCTTTTCATCGATCAGTGTTCATGGTTCATCGCGGTA 6548
 2018 euLeuGlnLeuGlyValPheAlaThrGlyAlaSerGluGlyLeuPhePhe 2034
 6549 GCGTACTAGCTTGCATTGAGTTTCACGAGGTGCATAAAGTAATGATCTTC 6598
 2035 GlyAspIleAsnGlnIleProPheIleAsnArg.....GluLysVa 2048
 6599 GGGGATAGCCGCGAGATTCACATATTGAAGGACGAATTTGGACAAGTG 6648
 2048 lPheArgMetAspCysAlaValPheValProLysLysGluSerValValT 2065
 6649 TTTGTATGGGATCTCGACAGGTTCTG...GACCTGCAGTGTTCGGGTTT 6695
 2065 yThrSerLysSerTyrArgCysProLeuAspValCysTyrLeuLeuSer 2081
 6696 ATGGTAATATTCTGACCGTGTTCATGGGATGTGTGCGCTTGGTTAAGC 6745
 2082 SerMet.....ThrValArgGlyThrGluLysCyst 2092
 6746 ACAGTGTATGGCAACCTAATCGCCACCGTGAAGGGTGAAGCGAA..... 6790


```
2092 rProGluLysValSerGlyLysAspLysProValValArgSerLeus 2109
6791 .....GGTAAGAGC.....AGCATGC 6806
2109 erLysArgProIleGlyThrThrAspValAlaGluIleAsnAlaAsp 2125
6807 GCATTAAACGAATAATTAGTCAGTCGACGATTAGTCCCGACGTTGGCTTCC 6856
2126 ValTyrLeuCysMetThrGlnLeuGluLysSerAspMetLysArgSerLe 2142
6857 ACGTTTCTGTGTCTTCACTCGGAGAGTTGGAAATCAGCAAGACACTT 6906
2142 uLysGlyLysGly...LysSerThrProValMetThrValHisGluAlaG 2158
6907 TATTCGCAAGGTTTGACTAACTTAACGTTCTAACGGTCATGAGCGGC 6956
2158 IncLysThrPheSerAspValValLeuPheArgThrLysLysAlaAsp 2174
6957 AAGGTGAGACGATGCGCTGTGAACCTTGTGCGACTTAAGTTTCAGGAG 7006
2175 AspSerLeuPheThrLysGlnProHisIleLeuValGlyLeuSerArgHi 2191
7007 GATGAACCTTTAAATCTATCAGGCACATAACCGTCGCTTCTCTGTCAC 7056
2191 sThrArgSerLeuValTyrAlaAlaLeuSerSerLysLeuAspAspLysV 2208
7057 CACCGACACTTAACTTAACCTTACTGCTCTGCTGAGGTGAGCGCCA 7106
2208 alcGlyThrTyrIleSerAspAlaSer 2216
7107 CTTGCGATGCCATCCAGAGGCTGCG 7132
```

seq_name: gb_vil:AF190581

seq_documentation_block:

```
LOCUS AF190581 15468 bp RNA VRL 04-JAN-2000
DEFINITION Beet yellows virus strain BYV-4, complete genome.
ACCESSION AF190581
VERSION AF190581.1 GI:6492367
KEYWORDS
SOURCE
ORGANISM
Beet yellows virus.
Beet yellows virus
Viruses: ssRNA positive-strand viruses, no DNA stage;
Closteroviridae; Closterovirus.
1 (bases 1 to 15468)
Peremyslov,V.V., Hagiwara,Y. and Dolja,V.V.
HSP70 homolog functions in cell-to-cell movement of a plant virus
Proc. Natl. Acad. Sci. U.S.A. 96 (26), 14771-14776 (1999)
20079557
10611288
2 (bases 1 to 15468)
Peremyslov,V.V., Hagiwara,Y., Alzhanova,D. and Dolja,V.V.
Direct Submission
Submitted (27-SEP-1999) Botany and Plant Pathology, Oregon State
University, 2082 Cordley Hall, Corvallis, OR 97331, USA
Location/Qualifiers
1. 15468
/organism="Beet yellows virus"
/strain="BYV-4"
/db_xref="taxon:12161"
/feature="Californian strain; RNA transcript derived from
this cDNA clone causes formation of red necrotic lesions
in Claytonia perfoliata and systemic infection in
Nicotiana benthamiana"
join(108..7997,7999..9381)
/feature="ORF1a/b; papain-like leader proteinase is capable
of autoproteolytic release from the rest of the
polyprotein, and is a replicational enhancer. RdRp is
presumably expressed via a +1 ribosomal frameshift"
/codon_start=1
/product="p348"
/protein_id="AA14300.1"
/db_xref="GI:6492368"
```

FEATURES

source

1. 15468

/organism="Beet yellows virus"

/strain="BYV-4"

/db_xref="taxon:12161"

/note="Californian strain; RNA transcript derived from this cDNA clone causes formation of red necrotic lesions in Claytonia perfoliata and systemic infection in Nicotiana benthamiana"

join(108..7997,7999..9381)

/note="ORF1a/b; papain-like leader proteinase is capable of autoproteolytic release from the rest of the polyprotein, and is a replicational enhancer. RdRp is presumably expressed via a +1 ribosomal frameshift"

/codon_start=1

/product="p348"

/protein_id="AA14300.1"

/db_xref="GI:6492368"

CDS

```
/translation="MAFLNVSAPVSCAFAPAFAPAHAGASPIVPDSFPCVPYRSDISH
RRLTLDLDFSVPRPLFLNARVHLRASTNDPLPSLPLGFHAETFYVLELNGSAPFSIPP
RHIDFVNRPSFVPTFTEPRLSRFLFALLDFELYSKPGPCVIAESFSTPSP
CLVNCVCAIPIHAEMESIRFPTKTLFAGRLQFHKKRYTKRPTLIITHEGLAKTS
ALGVTSKPNSRPIITVKSAGREKIYAYEISRKDFERSRRQTPRVRSHPKRIKAVE
PFPPEPKDKRKRASLPTKDEGFIETGLRFLPSETPKPEPLPKPREVPIPVVK
HAPVAVSKPVTFRPVATTCAGYVNAITQCSRPRNHPILRSASYTFGFKMPLQRF
MKERDYTVKRSKSVSSCSVPLEALTSILKNIIPRYSNRUKFYDHFIDGDDFEI
EYHPLRGKGLSVLLILPKGEAYCVVTAATPQYHAALTARGDRPRVGLLQYRGEGL
CYLAHALCCALQKTRFEEDFVGMPTKFVFAKRLTEKGLPSALPKVGRQVRSR
LFHCDVASAFSPFYSLPFTGGVEEAPEITSSLKHKAIESVYRVSHVNDLARS
VEKDLIDFKDKIKLSKFSKRSVTVTFVMEAVQGLTRAYQFNLFSHYSDHPAA
AGSRLENETLASMAKSSFSDIGCCPLFHKRGSTDYHVCPRIDYDMKDAQRRVRELQ
ARGVLENSRQGLVEAQRARVSVCPHTLGNVNVKSDVLIIMVOYDASLEIASAMVKE
SKVAYLTMTVQTELLDEREAFIDALGCDVVDTRDMVQYKFGSCYCHLSNLTNI
MLTPAFTSGNLSVEMYENRMGVNYKITRSAYSPSEIRGVKTYIRACRYEVGVKL
PRFDKTLTKTGLSGDYIYLDKAFVYRVDFVYVNCVSNVSKTFEWMVSYIKSSKRVV
ISGKVIHRDVIIDLKHSKFAVLAAGVRSRTTTFELAKNLNYTGDASCFTIRFL
PREMSRAYAEINRSFKLMSILSAGLDYFELDLNSLQHLLEYSEVSEVRVSTQNG
EVCNENRVLTEIIAADRKSIAQGLSSVPTQPRGLGRSSRSNFTLSVLYNL
VEEYGNLFESVDGAVRFLVKFTSDSPIFRVVRMFLDLAFAASPFVSVVSLCAWLR
EAVSAFSSVADRTVSESVKTVNTRFRFLNFSAKTLTKKFFRFLSALAKTVV
RKAKVILEAYWEVWFESILSDSGEYSAVEFCSSVITLTINSGRLLPCFSPASVITEV
LIDATKISIEVLLKQISPDSTASSALYRVLSKILSNFRMTGPHGLTPTVLLCGF
LPVFRKCVALCVGDMATYARFLEYGVDDLFELIGRSVNSIKNYLCVVAAGLVDSIVD
SVLKSQVAKERVLGPKSKIIKNFLANVRKAKVVTRTSSSTDLSEDEYFSCDESKPG
LRGSSRFTLSRLDIFNFLKSLVIENACFSAYERIENMKLIYFFLNSSEEAR
RLICAGDFDYLSDAFDEMLQAFQYYSDDDESVTYDGKPTLYSLANVSRFL
ETFCNGPKFFVKNYKALYSRLRLVLPWYDRNLSDSPGLKGGNEKALAKFLKTCV
ITACEGYSQICCLRLILCWCTPACGLVRLFYITSGYTRVLSRVVAVACPLVYRNE
LDGLSDGLTNMGVSVFRLFVALRRALSAYSNSALRRKIFEFIFGNHHPDVAIET
NEVAPELSPVDIDVDCDFGSDSESVDASIPRPLHGGSSRSNFTLSVLYVY
FKLAGRIPLFLRLNFVAYEVRRLASRLKTLFGLARLDFNLSLTVVYLQLYDS
VLNAFIDVELLNSGVNVLPSVWARGSLTKLAEVIVGSGFSLGCMCRVSDMC
SSSNAGCNFSPVTKGKFPVPPSSGSTASMYERLEALSDIRRHVLSTCRVSGDEE
PERKEVTEPGIETSDVPIRSHSOPLSGEGCSYSEDRENEANLPHVSKTYSE
RRGLTARNKRTLGVSEFLNAINTSNEQPRPIIVHSPESRALTNVRRPLYLQEL
ALFELSKLREYIDOLKAVANFNEBCLCDEDMFVLRAGOVSVGRNSRLPLKHFG
HEFCFRSGGLVPYDGTSRVDTFHTQTFVNSANALLSGLYSTRTFTTNLSANVLYE
APGGGKTTLIKVFCEFTFSKNSLILTANKSSREELAKVNRIVLDGDDTPLOTRDR
ILTIDSLMNNRGLTKVLYLDECFMHAGAAVACIEFTKCDLSAFLGDSRQIHVIDR
NELDTAVSLDLNRFVDDSESVYGVSVPCPDVCAMLSFEPKVTATTNLSAGQSSM
QVRETESVDDVEYSEFVYLMQSEKDKLLSKGRSSRVERPVTIVLTHEAQGETY
RKVNLVTRKFQEDDPPFRSENIHTVALSRHESLTVSLSSKRDADAIAQAKOLVD
AYRYPTSGGSTLDISVNPSTSDRSKCSASPEVINSELSVPTGTSDFDGDVS
EEMGTQVFGSADNVIRDSAPVNSKTDHPQVRSSIRSQAIIPKRKPSLQENLYSES
RNYNFTVCERESGPGQFOAMWMLERSFDELEKAVKRVSDVIAITEKGVRTWMSKRE
PSQJRALSSDLQKPLNLEEITTFKLVNKRDAKYKLDSSCLVKHHPHPPINPHKRVN
AIFSPCFDEFRNRVITCTNSNIVFTFTMTNSTLASIKEMLGSEHVVNVEIDPSKFD
KSQDAFIKSFERTLYSAPGDEDLDDVMQGEYTSNATTLQGLUSFSDNQRKSGASN
TWIGNSIVTLGILSMFYTNRFKALFVSGDDSLIFSESPIRNSADAMCTELGFETKFL
TPSVPYFCSEKFFVMTGHDFVPDYPKLLVKLGASKDEVDDEFLFEVTSFRLTKDL
VDERVIELLTHLVHSKYGESDGYAALCAIHCIRSNFSSEFKKLYPKYKGMVVHYGKL
KFVLRKFNCFEREXFDTAFGAYFLTYDET"
```

CDS

```
9445..9609
/note="small hydrophobic protein; ORF2"
/codon_start=1
/product="p6"
/protein_id="AA14301.1"
/db_xref="GI:6492369"
/translation="MDCVLRSYLLLAFLGLICLFLFCLVLFVFWVYKQILFRNTPPSN
EARNRSTVW"
9609..11405
/note="HSP70 protein homolog; ORF3; required for
cell-to-cell movement"
/codon_start=1
/product="Hsp70h"
/protein_id="AA14302.1"
/db_xref="GI:6492370"
/translation="MVVFGLDGTTFSVSCAVVGEELYLFKQORSAYIPTFVFLHSDT
QEVAFGDAEVLSDNDPSVGGFYRDLKRHWICDENYGDYLEKLPKHYKTELLKVAQS
SKSTVRLDCYSGTVFQNTATPLGIATFVKALISTASEAFKCOCTGVCISVPANYNCLQ
RSFTESCYNLSGPCVVMVNPBSAALSAKSRIKGTSPVLVYDFGGGTFDVSVTIAL
```


[illegible]

3454 TCAGCGGAGCGTGTAGCTCCGTTCCAAACGCAACCGAGCGGTCTGAGA 3503
:: : ||| :::::||||| ||:::|||||||
969 GlyGlySerLysAlaThr.....ValValSerSer 978
||:||||::: :||| ||:::|||||||
3504 GGGGGTAGTCGCGGAAGTGGAGTTCCTTATACCTTGGTAGAGGA 3553
||:||||::: :||| ||:::|||||||
978 rValSerLysGlyLeuValAlaAraGlyAlaAlaThrLysAlaPheSerG 995
||:||||::: :||| ||:::|||||||
3554 AGTCGGAATCTTTCTTCCTCGTCGGTATGTCGCGGTCTCTGTGA 3603
||:||||::: :||| ||:::|||||||
995 lyleThrSerPhePheSerThrGlySerLeuPheTyrAspArgGlyLeu 1011
:: : ||||| ::: :||| ||:::|||||||
3604 AGGTATTTAAAACTTTTCCGACTCTCCCATCTTTCGGTCTTCGGCATG 3653
||| ||||| ::: :||| ||:::|||||||
1012 ThrGluAspGluArgLeuAspAlaLeuValArgThrGluAsnAlaIleAs 1028
||| ||||| ::: :||| ||:::|||||||
3654 TTCTCGGAC.....TTGGCGGAACGAGCTTACCGCTTCGT 3688
||| ||||| ::: :||| ||:::|||||||
1028 nSerProValGlyIleLeuGluThrSerArgValAlaValSer..... 1042
||| ||||| ::: :||| ||:::|||||||
3689 TTCGGTGTCTTGTGTCGCGTGGTTCGCGAAGCTGTAAACACCTTTGTA 3738
||| ||||| ::: :||| ||:::|||||||
1043LysValValAlaGlyThrLysGluPheThrPser 1053
:::||||::: :||| ||:::|||||||
3739 CAAGCTGGGTTCGCCACAGACGGTGTGCGAAGCTGTAAACACCTTTGTA 3788
||| ||||| ::: :||| ||:::|||||||
1054 GluValSerLeuAsnAspPheThrThrPheValLeuArgAsnLysValLe 1070
:::||||::: :||| ||:::|||||||
3789 AACCTGACGGTAAAAAGGTTCTAACTTTATGTCGCAAAACCCCTGAC 3838
||| ||||| ::: :||| ||:::|||||||
1070 uileGlyIlePheValAlaSerLeuGlyAlaAlaProfileAla..... 1084
||| ||||| ::: :||| ||:::|||||||
3839 GAAAAATTTTTAGGTTTCTATCCGCTTCGCCCTAGCTAAACCCG 3888
||| ||||| ::: :||| ||:::|||||||
1085TrpLys...TyrArg 1088
||| ||||| ::: :||| ||:::|||||||
3889 TTGTCGAGGAAGCGAAGTGATTTTGGAAAGCTTACTGGGAAGTGTGTT 3938
||| ||||| ::: :||| ||:::|||||||
1089 ArgGlyIleAlaAlaAsnAlaArgArgTyrAlaGlySerSerTyrGluTh 1105
:::||||::: :||| ||:::|||||||
3939 GAATCGATTTCTTCGATAGTGGAGAGTAT.....AGTGTGTTTGAAT 3982
||| ||||| ::: :||| ||:::|||||||
1105 rLeuSerSerLeuSerSerGlnAlaAlaGlyLeuArgGlyLeuThrS 1122
||| ||||| ::: :||| ||:::|||||||
3983 TTGTCAGCAGCGTTGTG.....ATCACAC 4005
||| ||||| ::: :||| ||:::|||||||
1122 erSerThrValSerGlySerLeuValValArgArgGlyPheSerSer 1138
||| ||||| ::: :||| ||:::|||||||
4006 TGTAAACGAATTCGCGTCCGCTGCTA.....CCTGGGTTTGTAGTCCA 4046
||| ||||| ::: :||| ||:::|||||||
1139 AlaValThrValThrArg.....AlaThrValAlaLysAr 1150
||| ||||| ::: :||| ||:::|||||||
4047 TCCGCTGTCTATTCATTAAGTACTGCTGACCTAGCCAGCAAAATATCCAT 4096
||| ||||| ::: :||| ||:::|||||||
1150 gGlnValProLeuAlaLeuLeuSer.PheSerThrSerTyrAlaIleSer 1166
||| ||||| ::: :||| ||:::|||||||
4097 TGAAGTTCTCTTAAACAANAATTTCTCTGTGCATCTGCACAGCTTCTTCGG 4146
||| ||||| ::: :||| ||:::|||||||
1167 Gly.....CysSerMetLeuGlyIleThr 1174
||| ||||| ::: :||| ||:::|||||||
4147 CCTTATATCGCCGCGTCTTAACGAGATTCGTGCAATTTTCGCACAATG 4196
||| ||||| ::: :||| ||:::|||||||
1174 palAHisAlaLeuProArgHisLeuMetPhePheGlyLeuGly.... 1189
||| ||||| ::: :||| ||:::|||||||
4197 GCGCAACACG...GGATCTTCACTAAAGCTTTTTACTCTCGCGGGTTCT 4243
||| ||||| ::: :||| ||:::|||||||
1190ThrLeuLeuGlyAlaArg.....AlaSerAlaAsnThrTrp 1201
||| ||||| ::: :||| ||:::|||||||
4244 ACCGGTGTGTGTAGAAAGTGGGTGCTCTGTGCTTCTCGTGCAATGG 4293
||| ||||| ::: :||| ||:::|||||||
1202 ...LysPheGlyGlyPheSerAsnAsnTrpCysAlaValProGluValVa 1217
::: ||| ||||| ::: :||| ||:::|||||||

4294	CGACGTACGCCCGGTTCGCAATAC.....GGGTCGACGACTCTCTT	4330
1217	lTrpArgGlyLysSerValSerSerLeu.....LeuLeuProIle	1231
4337	CTTTTGGGAGGTCCTCGTGAACCTCGATTGAACAACACTACTATGTGTGGTGG	4386
1231	hrLeuGlyVal.....	1234
4387	CTGCGGGTGTGTGGACTCCATTGTCGATTCGGTGTGTTTGAACAACTTTCC	4436
1235SerLeullelleArgG1	1240
4437	GGTGTGCCAAAGACGAGTCTTGTTTAAATCAAAATCATATAAAAA	4486
1240	yLeuLeuAsn.....	1243
4487	TTTTTTAAACGTCTTAGGAAGGCAAGTCGTTCACAGTACTCTTCTTCCA	4536
1244AspThrIlePro	1247
4537	GCACGGACTTGTGGAAGACGAATATTTTCGTGCGACGAAGCAACACC	4586
1248	GlnLeu.....	1249
4587	GGCTTCAGAGCGGTCTCTAGGTTACGGTTTCGGCATTTGCTTGACAT	4636
1249	1249
4637	CTTTTTCAAATTCCTGAAAGTTCAAAGCTCGTCATAGAGAATGCGTGCT	4686
1250AlatyrValProProValGluGlyArgAsnValTyr.....	1261
4687	TTTCCTTCACGAAGGATTGAAAGGACATGAACATATACTTTTTCCT	4736
1262AspGluThrLeuArgTyrTyrArg.....As	1270
4737	TTAAATCTCTCGAAGAAGAAGCTCGCGCTTAATTCGGTCGCGGGAGA	4786
1270	pPheAspTyr.....AspGluGlyValaGlyProSerG	1281
4787	CTTCGACTACTTAAAGCGATAGCGGTTTCGACGAAGATGAATGTTACGGC	4836
1281	lyThrClnHisGluAlaValProGlyAspAspAsn.....Asp	1293
4837	ANGCTTTCGAACAATACTATTCCAGTCGACGACGAGAGTGTAACCTACGAT	4886
1294	GlySerThrSerSerValSerSerTyrAspValValThrAsnValArgAs	1310
4887	CGAAACCCACCAAGTTCCTCGGTAGTTACTTTGAACGTGAGTAGA...AGGTT	4933
1310	pValGlyIleSerThrAsnGly.....GluValThrGlyC	1322
4934	CTTGGAACAGTCTGTAAACGCCCGCAAGTTTTCGTCAAAAGTTTCGAATT	4983
1322	luGluGluThrHisSerProArgSerValGln...TyrThrTyrValGlu	1337
4984	ACTTTAAAGCGCTGTATAGTCGTTTTCGTCGGTCTACCGTGGGTAGAC	5033
1338	GluGluValAlaProSerAlaAlaValaLacIuArgGlnGlyAspProse	1354
5034	AGAAATCTTCTGTGACTCG.....CCAGGACTTAAAGG	5065
1354	rclySerGlyThrAlaAspAlaMetAlaPheVal.....	1365
5066	AGGTAAACGAGAAAGCTCTTCTTGCGAAGTTCCTTTAAACCTGCGTCATAA	5115
1366GluSerValLysLys.....	1370
5116	CCGCTTGCAATGTGTGCGCAGATATGCTGCCCTGCCTTATTTCGCTTA	5165
1371GlyValAspValPheHisGlnGlnse	1380
5166	TGTTGGGCGGACACGAGCGGTGGTTAGTTAGTTATTTTACATAACTTAA	5215

1380 rSerGlyGIuThrAlaArgGluValGluluaValAspGlyLysGlyLeuLeuP 1397
||||| : : : ||| : : : :
5216 TTCGGTACTCGCGTTGTTCGCGTGTTGTCTCGCGTGCCTGTATGCC 5265
roGLuSerValValGlyGluLaPro..... 1405
|| ||| : : : |||
5266 CTCCTTTTGGTCAGAAACGAACGATTGGATGCTTGAAGTGGATTAAACAAC 5315
1406 ThrGlnGluArqGlyAr 1411
5316 ATGGCGCTTTCGGTTTTCGTCGCTGTTTCGTCGCCCTTCGGCGCGCGCT 5365
1411 gAlaLaAspGlyAsnThraLa..... 1418
: : : : : : : : : : : : :
5366 TTCGGGTACTCTAATTCCGCCCTACGACGACGAGAAGATTTTCGAATTCATTT 5415
1419 GlnThrAlaVal..... AsnGlu 1424
5416 TCGGAACATTCACCACCCCTTTTGATGTCGCAGTAATACTGAACAACAAACAA 5465
1425 GlyAspArqGluProValGlnSerSerLeuValSerSerProGlnAlaAs 1441
||||| : : : : : : : : : :
5466 GTTGGCGCGGAACCTCTT..... TCACGGGAAGTGA 5497
1441 pileProLys..... ValThrGlnSerGluValHislaGlnL 1454
|||| : : : : : : : : : :
5498 CATTGATGTCGACTCGGACTTTGGTTCCGATTACAGATCGGTTCTTCAG 5547
1454 ysGluValLysGlnGluValProLeuAlaThrValSerGlyAlaThrPro 1470
||||| : : : : : : : : : :
5548 ATGAAGTC... CGCTCGATTCCCCGTCACGCTTACACGGTGGGAGT... 5591
1471 IleValAspGluLysProAlaProSerValThrThrArgGlyValLysII 1487
5592 AGCGCTCTTCCAACCTTCTAACCTCTCTCGTGAAGT 5629
1487 eileAspLys..... GlyLysAla..... 1493
: : : : : : : : : : : : :
5630 TGTTTTTAAGTTGGCTGGCGCATTCGCGTTTCGTTTCGTTACGCA 5679
1494 ValalaHisValalaglutylusGlnValGlnValGluInPro 1508
||||| : : : : : : : : : :
5680 AATTTCGCGCTACTTTGTAACACGCGCTGGCT..... 5714
1509 LysGlnArqSerLeuThrIleAsnGluClyLysalaglyLysGlnLeuCy 1525
5715 ...TCAAAAGGTGTCAACACGTTTCATCGGCTCGGCC... AGGCTGTTTGA 5758
1525 sMetPheArqThrCysSerCysGlyValGlnLeuAspValTyAsnGluA 1542
||| ||| : : : : : : : : : :
5759 TAATTCTCATCTCCTCGTAGTTTACCTTCTCCAGAGTACGATTTCAG 5808
1542 laThrIleAlaThrArgPheSerAsnAlaPheThrPheValAspAsnLeu 1558
: : : : : : : : : : : : : : : :
5809 TGCTG..... AACGCATTTATAGACGTTGAGCTAGTC 5840
1559 LysGly.. ArgSerAlaValPhePheSerLysLeuGlyGlu 1571
5841 CTACTTAACTCGGTAGCGTAAGTGTCTACCTTACTTGTCTTCTGGGTTCAG 5890
1572 GlyTyThrTyraSn... GlyClySerHis.. ValserSerClyTyrrProA 1587
||| : : : : : : : : : : : : : : : :
5891 GGGATCTCTAACGAAGTTAGCGGAAGTATCGTTGGTTTCAGGCTTCGCTT 5940
1587 rgAlaLeuGluAspileuthrAlaIleLysTyrrProSerValPheAsp 1603
||| : : : : : : : : : : : : : : : :
5941 CTTTCTAGGAAGGATGTC..... 5960
1604 HisCysLeuValGlnLysTyrrLysMetGlyClyGlyValProPheHisAl 1620
||| ||| : : : : : : : : : :
5961 ...TCTCGTGTCTCCACTGG..... 5978

```

1620 aAspAspGluGluCysTyrProSerAsnProIleLeuThrValAsnL 1637
      ||| ||| ||| ||| |||
5979 .....TGCTCCCTCGCTTCGAAC..... 5996
1637 euValGlyLysAlaAsnPheSerThrLysCysArgLysGlyLysVal 1653
      ||| ||| ||| ||| |||
5997 .....GCCGGTTGTAACCTTTATGAGTCCGGTTCGTACGAAGAAG... 6038
1654 MetValIleAsnValAlaSerGlyAspTyrPheLeuMetProCysGlyPh 1670
      ||| ||| ||| ||| |||
6039 TTCGTTCCCTTCGCTTCCTCGGTTCAACCGCTTCGATG.....TA 6079
1670 eGlnArgThrHisLeuHisSerValAsnSerIleAspGluGlyArgIles 1687
      ||| ||| ||| ||| |||
6080 TGAACGT.....CTCGAAGCTCTCGAGACGGATATCCGCGAACACGTGC 6123
1687 erLeuThrPheArgAlaThrArgArgValPheGlyValGlyArgMetLeu 1703
      ||| ||| ||| ||| |||
6124 TTTCACAGTGTCTG.....GTAGCA..... 6143
1704 GlnLeuAlaGlyGlyValSerAspGluLys..... 1713
      ||| ||| ||| ||| |||
6144 .....ACGCGACGAAGAAGAAAGGCCGGAAGAAGT 6175
1714 ....SerProGlyValProAsnGlnGlnProGlnSerGlnGlyAlaThrA 1729
      ||| ||| ||| ||| |||
6176 GACAGAACCAAGGAATTAACACTTCTCGAAGTGAATGTTCTGATTCGATCGT 6225
1729 rgThrIleThrProLysSerGlyGly..... 1737
      ||| ||| ||| ||| |||
6226 CACACTCGCAACCTTATCTGAGGTTGAATGTTCTGATTCGATTCGATCGT 6275
1738 ..... 1741
      ||| ||| ||| ||| |||
6276 GAAGAGAATGAACGAGCAACCTGTTACCGCACGTTAGCAAAATCGTCAG 6325
1741 rGluGlySerGlyArgGluValLysGlyArgSer..ThrTyrSerIleThr 1757
      ||| ||| ||| ||| |||
6326 CGAACGAAGGGGTTGGAGACCCCGCTCGAAACAAACGACTCTACGTT 6375
1757 pCys.GluGlnAspTyrValArgLysCysGluTrpLeuArgAlaAspAsn 1773
      ||| ||| ||| ||| |||
6376 GTGTAGC.....GAGTTTCTCAACGCTATTAAT 6404
1774 .....ProValMetAlaLeuGluProAspTyrTh 1783
      ||| ||| ||| ||| |||
6405 ACTAGCAATGAGCAACCTAGCGCGAATAATCGTT.....GACCACCTC 6445
1783 rProMetThrPheGluValValLysThrGlyThrSerGluAspAlaValV 1800
      ||| ||| ||| ||| |||
6446 TCCTGAATCT.....GCCGGTTGACCAACTCCGTCGAGGGAAT 6483
1800 alGluTyrLeuLysTyrLeuAlaIle.....GlyIleGlu 1811
      ||| ||| ||| ||| |||
6484 TCTATTACCTCCAGGAACCTCGCTCTTTTTCAGTTGAGTTCCAAACTCGT 6533
1812 ArgThrTyrArgAlaLeuLeuMetAlaArg..... 1821
      ||| ||| ||| ||| |||
6534 GAGTATTACGATCAATTTGAAGTTGCGAATTTTAACAGACAAGAGTGT 6583
1822 .....AsnIleAlaValThrAlaGluGlyValL 1832
      ||| ||| ||| ||| |||
6584 GTCCGCAAAAGACAGACATGTTTCTTACGAGCGCGGAAGGTGTAG 6633
1832 euLysValProAsnGlnValTyrGluSerLeuProGlyPheHisValTyr 1848
      ||| ||| ||| ||| |||
6634 TTTCGGCGCAGAAC.....TCGAGGTTGCTCTTAAAGCAATTC... 6671
1849 LysSerGlyThrAspLeuIlePheHisSerThrGlnAspGlyLeuArgVa 1865
      ||| ||| ||| ||| |||
6672 ...AAGGGTCACCAATTTTCTTTTCGCTCT.....GGAGGGTTG..... 6707
1865 lArgAspLeuProTyrValLeuIleAlaGluLysGlyIlePheThrLysG 1882

```

```

6708 .....GTCCCTTAC..... 6716
1882 lYAspValAspAlaValAlaLeuGlyAspAsnLeuPheValCys 1898
      ||| ||| ||| ||| |||
6717 .....GACGCTACCAGCAGAGTG 6734
1899 AspAspIleLeuValPheHisAspAlaIleAsnLeuIleGlyAlaLeuLy 1915
      ||| ||| ||| ||| |||
6735 GACACCATT.....TTTCACACGCAAAACGAATTTCTGTTCCCGG...AA 6775
1915 sValAlaArgCysGlyMetValGlyGluSerPheLysSerPheGluTyrL 1932
      ||| ||| ||| ||| |||
6776 CGCGCTTCTTCGGGCTATCTC.....TCTATAGAACTTTTCACTTTCA 6819
1932 ys.....CystYrAsnAlaProProGlyGly 1941
      ||| ||| ||| ||| |||
6820 CTAATTTGAGCGCTAAGCTACTGCTGTACGAAGCTCTCCAGGTGGTGA 6869
1942 LysThrThrThrLeuValAspGluPheValLys.....SerProAsnSe 1956
      ||| ||| ||| ||| |||
6870 AAGACCACGACTTTGATAAAGGTTTGTGAAACTTTTCAAAAGTTAA 6919
1956 rThrAlaThrIleThrAlaAsnValGlySerSerGluAspIle..... 1970
      ||| ||| ||| ||| |||
6920 TTCCTGATTTTACGGCGAATAGAGTTCCGGAAGAATAACTTTCGCA 6969
1971 .....AsnMetAlaValLysLysArgAspProAsnLeuGluGlyLeu 1984
      ||| ||| ||| ||| |||
6970 AGGTGAATCGCATCGTACTTGCACGAAGCGCATCGCTCTTCAGACCGGT 7019
1985 AsnSerAlaThrThrValAsnSerArgValValAsnPheIleValArgG 2001
      ||| ||| ||| ||| |||
7020 GACAGGATTTTAACTATCGATTTCTTATTAATGAAC.....AACAGAGG 7063
2001 yMetTyrLysArgValLeu...ValAspGluValHisMetMethHisGln 2017
      ||| ||| ||| ||| |||
7064 TTTGACGTGTAAAGTTTGTACCTCGATGAGTGTTCATGTTTCCACGCTG 7113
2017 lYLeuLeuGlnLeuGlyValPheAlaThrGlyAlaSerGluGlyLeuPhe 2033
      ||| ||| ||| ||| |||
7114 GAGCTGCTGTAGCTTGCATCGAATTCACCAATGCGGATTCAGCCATCTTA 7163
2034 PheClyAspIleAsnGlnIleProPheIleAsnArgGluLysValPheAr 2050
      ||| ||| ||| ||| |||
7164 TTTGGAGACAGACGCAATTCATACATAGACCGTAACGAA..... 7205
2050 gMetAspCysAlaVal.....PheValProLysLysG 2061
      ||| ||| ||| ||| |||
7206 .TTGGATACCTGCTGTTCTTTCAGATTTGAACCGGTTTGTCT...GATGATG 7251
2061 luSerValValTyrThrSerLysSerTyrArgCysProLeuAspValCys 2077
      ||| ||| ||| ||| |||
7252 AATCGAGAGTATATGGTGAAGTCTCATACAGGTCTCTTGGAGGTTTGT 7301
2078 TyrLeuLeuSerSerMetThrValArgGlyThrGluLysCysTyrProG 2094
      ||| ||| ||| ||| |||
7302 GCTTGTGTCAACTTC.....TACCCGAA 7327
2094 uLysValVal.....SerGlyLysAspLysProValV 2105
      ||| ||| ||| ||| |||
7328 AACTGTGCCCACTACCAACTTGTGTTTCAGCGCGTCAATCT..... 7367
2105 aArgSerLeuSerLysArgProIleGlyThrThrAspAspValAlaGlu 2121
      ||| ||| ||| ||| |||
7368 .....TCGATGCAAGTACGCGAGATTCGAACCGTAGACGAGCTGAATAT 7412
2122 lLeAsnAlaAspValTyrLeuCysMetThrGlnLeuGluLysSerAspMe 2138
      ||| ||| ||| ||| |||
7413 TCCAGTGAATTCGCTACTTACTATGACTATGTTACAGTCACAGAAGAAATCT 7462
2138 t.....LysArgSerLeuLysGlyLysGlyLysGluThrP 2150
      ||| ||| ||| ||| |||

```


1217 lTrpArgGlyLysSerValSerLeu.....LeuLeuProIleT 1231

4337 CTTTTTTGGGAAGTCCTCGTAACCTCGATTTAAGAACAACCTAACCTATGTGTGGTTG 4386

1231 hrlLeuGlyVal..... 1234

4387 GCGCCGGGTGTGGACTCCATTGTGCAGATCCGTTGTTTGAACAATTTC 4436

1235SerLeuIlellelAargGl 1240

4437 GGTGTGCCCAAGAGCAGAGTGCTTGGTTTAAATCATAAAATCATATAAAAA 4486

1240 yLeuLeuasn..... 1243

4487 TTTTTTAAACGTCCTTAGGAAGCGCAAAGTCGTTACACGACTTCTTCCA 4536

1244AspThrIlePro 1247

4537 GCACGGACTTGTGGAAGACGAGNATATTTTCGTGCGACGAAGCAAAACC 4586

1248 GluLeu..... 1249

4587 GCCTTGAGAGCGGTTCTCTAGGTTCACGCTTTCGCGATTGCTTGACAT 4636

1249 1249

4637 CTTTTTCAAATTCCTGAAAGTTCAAAGCTCGTCATAGAGAATCGTGCT 4686

1250AlatYrValProValGluGlyAArgasnValTyr..... 1261

4687 TTCTGCTTACGAAAGGATTGAAAGGAACATGAACATATATCTTTTTCCT 4736

1262AspGluThrLeuAryTyrArg.....As 1270

4737 TTAATTCCTCGAAGAGAAGCTCGCGCTTAATTCGTGCGCGGAGA 4786

1270 pHeAspTyr.....AspGluGlyAlaGlyProSerG 1281

4787 CTTGCACTACTTAAGCGATAGCGCTTCGACCAAGATGAATGTTACGGC 4836

1281 LyThrCluHisLuuAlaValProGlyAspaspasn.....Asp 1293

4837 AGCTTTGGAACAATFACCTATTCAGTAGCAGCAGAGAGTGTAACCTACGAT 4886

1294 GlySerThrSerSerValSerSerTyrAspValValThrAsnValArqas 1310

4887 GGAAACCCACAGTCTCCGTAGTTACTTGAACGTGAGTAGA...AGTT 4933

1310 pValGlyIleSerThrAsnGly.....GluValThrGlyG 1322

4934 CTTGGAAAGCTTCTGTAACGCCGCGAAGTTTTTCGTCAAAGTTTCGAATT 4983

1322 LuGluThrHisSerProArgSerValGln...TyrThrTyrrValGlu 1337

4984 ACTTTAAAGCGCTGTATAGTCGTTGTCCGTGTCTACCGTGGGTAGAC 5033

1338 GluGluValAlaProSerAlaAlaValAlaGluArgGlnGlyAspProse 1354

5034 AGAATCTTCTGACTCG.....CCAGACTTAAAGG 5065

1354 rGlySerGlyThrAlaAspAlaMetaIaPheVal..... 1365

5066 AGGTAAACGAGAAAGCTCTCTTCGCGAAGTTCCTTAAACCTCGCTCATAA 5115

1366GluSerValyllys..... 1370

5116 CGCTTGCGAATGTGTGTCGACAGATATGCTGCCTGCCTTATTTCGCTTA 5165

1371GlyValAspValPheHisGlnInSe 1380

5166 TGTTGGGGACACCACTGCTGTTAGTTAGTTATTTTACATAACTTAA 5215

1380 rSerGlyGluThrAlaArqGluValGluValAspGlyLySGlyLeuLeup 1397

5216		:::	:::		:::		1
	TTCCGGTACTCGGGTGTTGTGCGCGTGTGTGGTCGGGTGCGTGTATGCC	5265					
1397	roGluSerValValIGlyValAlaPro						
		:::	:::				
5266	CTCTTTGGTCAGAAACGAACCTGGATGCGTTGAGTGTGATGATTAACCAAC	5315					
1406	ThrGlnGluArgGlyAla	1411
	1411
5316	ATGGGGCTTTCCGTTTTTTCGTGCGCTTGTTCGTGCGCCCTTCGGCGCGCGCT	5365					
1411	gAlaAlaAspGlyAsnThrAla						
		:::	:::				
5366	TTGGGGCTACTCTAAATTCGCGCCCTACGACGGAAGATTTCGAATTCAATTCATTT	5415					
1419	GlnThrAlaVal	1424
	1424
5416	TCGGAACATTCACCAACCTTTGATGTGCGAGTAATCGAAACAAACGAA	5465					
1425	GlyAspArgGluProValGlnSerSerLeuValSerSerProGlnAlaAs	1441					
		:::	:::				
5466	GTTCGCCCGCAACCTCTT					TCACCGGAAGTCGA	5497
1441	pIleProLys					ValThrGlnSerGluValHisAlaGlnL	1454
		:::	:::			1454
5498	CATTGATGTCGACTCGCACTTGGTTCGATTACAGATCGGTTCTTCAG	5547					
1454	ysGluValLysGlnGluValProLeuAlaThrValSerGlyAlaThrPro	1470					
		:::	:::				
5548	ATGAAGTC...GGTCGATTCCCGCTCCAGGCTTACACGGTGGGAGT	5591					
1471	IleValAspGluLysProAlaProSerValThrThrArgGlyValLysI	1487					
	1487
5592	1487
1487	elleAspLys					GlyLysAla	1493
		:::	:::			1493
5630	TGTTTTTAAGTTGGCTGGCGCACTTCGCGGTTTGCCTTTCCGTTTACGCA	5679					
1494	ValAlaHisValAlaGluLysLysGlnValGlnValGluGlnPro	1508
		:::	:::			1508
5680	ATTTTGTGGCGTACTTGTGTTGAACGACGGCTGGCT	5714					
1509	LysGlnArgSerLeuThrIleAsnGluGlyLysAlaGlyLysGlnLeuCy	1525					
		:::	:::			1525
5715	...TCAAAAGTTGAACAGCTTCATCGCTCGCC	5758				AGGCTGTTTGA	5758
1525	sMetPheArgThrCysSerCysGlyValGlnLeuAspValTrpAsnGluA	1542					
		:::	:::			1542
5759	TAATTTCTCACTCATTCCGTAGTTTACCTTCTCCAAGATGACGATTACAG	5808					
1542	laThrIleAlaThrArgPheSerAsnAlaPheThrPheValAspAsnLeu	1558					
		:::	:::			1558
5809	TGCTG					5840
1559	LysGly	1571
		:::	:::			ArgSerAlaValPhePheSerLysLeuGlyGlu	1571
5841	CTACTTAACTCGGGTACGGTAATGTGCTACCTTTAGTTTCTGGGTACAG	5890					
1572	GlyTrpThrTrpAsn...GlyGlySerHis	1587				ValSerSerGlyTrpProA	1587
		:::	:::			1587
5891	GGGATCTTAACGAAGTTAGCGGAAGTGCATTGGTTGGTTCAGGCTTCGCTT	5940					
1587	rgAlaLeuGluAspIleLeuThrAlaIleLysTrpProSerValPheAsp	1603					
		:::	:::			1603
5941	CGTTTCTAGGAAGGATGTC	5960					
1604	HisCysLeuValGlnLysTrpLysMetGlyGlyValProPheHisAl	1620					
		:::	:::			1620
5961	...TGTCGTGTGCCGACTGG	5978					
1620	aAspAspGluGluCysTrpProSerAspAsnProIleLeuThrValAsnL	1637					
		:::	:::			1637

```

5979 .....TCGCTCCTCGCTTCGAAAC.....5990
1637 euValGlyLysAlaAsnPhcSerThrLysCysArgLysGlyGlyLysVal1653
    ::::|||||:::|||||
5997 .....GCCGGTTGTAACATTATGAGTCCGGTTCGTACGAAGGGAAG...6038
1654 MetValIleAsnValAsnSerGlyAspTyrPheLeuMetProCysGlyPh1670
    ::::|||||:|||||:::|||||
6039 TTTCGTTCCCCCTTCGTCITTCGGTTCACCGCTTCGATG.....TA6079
1670 eClnArgThrHisLeuLeuHisSerValAsnSerIleAspGluClyArgIleS1687
    ::::|||||:::|||||:::|||||
6080 TGAACGT.....CTCGAAGCTCTCGAGAGCGATATCCGCGCAACACGTGC6123
1687 erLeuThrPheArgAlaThrArgValPheGlyValGlyArgMetLeu1703
    ||| ||| ||| ||| ||| ||| ||| |||
6124 TTTCACAGTCTCGT.....GTAGGA.....6143
1704 GlnLeuAlaGlyGlyValSerAspGluLys.....1713
    ||||| |||||:::
6144 .....AGCAGCAACAAGAAAGAAAGGCCGAAAGAAGT6175
1714 ....SerProGlyValProAsnGlnProClnSerGlnGlyAlaThrA1729
    ::::|||||:::|||||:::|||||
6176 GACAGACAGCAAGATTGAACATACTCTCGAAGATGTTGTTCCCATTCGTT6225
1729 rgThrIleThrProLysSerGlyGly.....1737
    ||| ||||| ||||| ||||| |||||
6226 CACACTCGCAACCCCTATCTCGAGGTGAATCTTCGTATTCGTAAGATCGT6275
1738 .....LysAlaLeuSe1741
    ||||| |||||:::
6276 GAAGAGAATGAACGAGCAACCTGTTACCGCACGTTAGCAAAATCGTCAG6325
1741 rGluGlySerGlyArgGluValLysGlyArgSer..ThrTyrSerIleThr1757
    ::::|||||:::|||||:::|||||
6326 CGAAGCAAGGGGTTTGGAGACGGCCGCTCGAAACAACAGTACTACGCTG6375
1757 pCys_GluClnAspTyrValArgLysCysGluTyrLeuArgAlaAspAsn1773
    |||||::: |||||::: |||||::: |||||
6376 GTGTAAAGC.....GAGTTTCTCAAGCTATTAAAT6404
1774 .....ProValMetAlaLeuGluProAspTyrTh1783
    |||||::: |||||::: |||||:::
6405 ACTAGCAATGAGCAACCTAGGCGGATATCGTT.....GACCACCT6445
1783 rPrometThrPheGluValLysThrGlyThrSerGluAspAlaValV1800
    :|||::: |||||::: |||||:::
6446 TCCTGAATCT.....CGCGGTTTGACCAACTCCGCTGAGGGAAT6483
1800 aGluTyrLeuLysTyrLeuAlaIle.....GlyIleGlu1811
    |||||::: |||||::: |||||::: |||||
6484 TCTATTACCTCCGGAACCTCGCTCTTTTGGATTGAGTTGCAAACTTCGT6533
1812 ArgThrTyrArgAlaLeuMetAlaArg.....1821
    :||| ||| ||| ||| ||| ||| ||| |||
6534 GAGTATTACGATCAATTGAGGTTCCGAATTTTAACAGACAAGAGTCTTT6583
1822 .....AsnIleAlaValThrAlaGluClyValL1832
    :|||::: |||||::: |||||::: |||||:::
6584 GTGCGCAAAAGACGAAGACATGTTGTTCTACGAGCGGACAAGGTGAG6633
1832 euLysValProAsnGlnValTyrGluSerLeuProGlyPheHis.....1846
    ::::|||||:::|||||:::|||||
6634 TTTCGGCGCAGAAAC.....TCGAGGTTCGCTCTTAAGCATTTCAAG6674
1847 .....ValTyrLysSerGlyThrAspLeuIlePheHisSerTh1859
    ::::|||||:::|||||::: |||||
6675 GATCAGCAATTTTGTTCGCTCTCGA.....6701
1859 rGlnAspGlyLeuArgValArgAspLeuProTyrValLeuIleAlaGluL1876
    ||||| |||||::: |||||::: |||||
6702 .....GGGTG.....GTCCCTTAC.....6716

```

[illegible]


```
SKSTVKLDCVSGTVPQNATLPLGLIATFVKALISTASEAFKQCTGTCVSPANVNCIQ
RSTFESCVNLSGYPVCMVNEPSSAAALSACSRIKGTATSPVLVYDFGGTDFSVISAL
NTTFVVRASGCDNMNIGDIDKAFVEHLYNKQAQLPVNYKIDISFLKESLKKVSLNF
PVTVEQGVVDVLNVNVEAEAAAPFVERTIKVKEYEYKCSMRLEPNVAKLLMW
GGSSYLPGLLSRLSSIPIFVDECLVLPDARAAGVAGGCALYSACLRNDSPLMLVDCAAHN
LSTSKKCEIVCPAGSPPIPTGVRTVNMGTGNASAVYSAALFEGDFVCRKLKRIF
QGDVLGVNAGTATVPTLLEINVSSVGTISFSLVGPVGKLLGGNAAVDFFSY
FLGERVADVJHKNISDKVLIHALTYQPFQRKLTLDGDKALFLKRLTADYRREARKFS
SYDDAVLNSSELLELRIPLKLRGSRVEKLDV"
11310..112971
/notes-"putative heat shock 90-related protein"
/codon_start-1
/protein_id-"CAA51866.1"
/db_xref-"GI:405629"
/db_xref-"SPTREMBL:Q08541"
/translation-"MTTRESTPANYWYGGELFRFFGGQEWKNLMSEAAVSRRPRYSDD
FRSDGVLSRKFTGSGTSEFVREPSLLITPKTYEVCHLGVAMELALGNRLSD
YNNSEINVDVKTGCKFNQSVTEFVKKNGNVAEPLSVLVEHCWSLNSGCELINPKD
TKRFVSLIFGKDLAESTDEAIVSSSYLDYLSHCLNLYETCNLSSNSGSKSLYDEFLL
LIDYLSNDSLEYSPSDNPLVAGIILYDMCFEYNTLKYLNIESFDCFLSLYLPLL
SEVFSMMWRPAPDRLFLFELDAEALLKVPYTNMHDSTFLYKNKRLYLSYFEDDSN
ELTKVVDLSLTNDNPELKLQARWGVGHYGVFRTAOTRKVRDAEYKLPALGPV
INNSGVFEFEELQKMPISVRRRECGSLSEAFSVPKBFGVGFPIRLNVPVKS
YLNVDYRHKVRVGLTQDELTILSNEFDSVAECCREVALQARRAORGEKPTQGMKG
TKNEISPHARSSIRVKRKNDSLLNLMKDVGAQRKRLNPLHRRK"
12919..13569
/codon_start-1
/product-"coat protein homolog"
/protein_id-"CAA51867.1"
/db_xref-"GI:405630"
/db_xref-"SPTREMBL:Q08538"
/translation-"MLAPEARGDLIIHFTENTRDMETFFNSYDLAEYSEVNPKNLNRK
ETDELQVIREKESLVIDEFDVFKHLAPALIRAAINITTSRVKNVYVGAHEYITGGK
FLVKDAMVPLIKCKMKKPNPVRTEFCATFEDIV IARSPLKFLNRTICKRGIP
SGVEFLGADFLTATSVCLNDHKAIVLOASRAAIDRAVSSVDGKIVSLFDLGRLS"
13640..14254
/codon_start-1
/product-"coat protein"
/protein_id-"CAA51868.1"
/db_xref-"GI:405631"
/db_xref-"SPTREMBL:Q08531"
/translation-"MGAEPISAIATFENVSLADQTLHGECDDKLRKNFEELKLG
VPDNLGIALGLCLYSATIGTSKNKNVQPTSFIRASFGGKELYLTHGELNSFLGS
OKLEKPKNLCFCRTFQKDYISLRKEYRKLPIIARANRHGLPAEDHYLAADFIST
STELTDIQSRILLARENATHFESESPTVSLKQLGRGLGTGR"
14251..14793
/notes-"orf7"
/codon_start-1
/protein_id-"CAA51869.1"
/db_xref-"GI:405632"
/db_xref-"SPTREMBL:Q08544"
/translation-"MTSSVBLAQTPLFRVLLKGVFVYVAFETEESEAEPLPVY
LHDFELNINRKGKIEASYVDFMCMTRIKPSSVSYTRVSEKSEDFSLPCSKTGS
KVLNRKVTTFENGVLVFGMYGLEQRVSSDYIWNFENVFVGAHCGLTLYCLNCELDK
SGGELELTFSKNEVLLRW"
14768..15301
/notes-"orf8"
/codon_start-1
/protein_id-"CAA51870.1"
/db_xref-"GI:405633"
/db_xref-"SPTREMBL:Q08545"
/translation-"MKFFLDGSETRALSRSESLLRVRKELGTNSQOSESECVDFEN
ELASFHLLVTVHEHMEQHPNQSKLRVPSIGEMLKEIRAFILKRVVYTPMHKETA
SDTLNAAFLECYCRITGLIARCDALRERMRKVSVLFFHSELLEKFEVTENMFSYTELLK
LNLSLRVSSOILGMAI"
15299..15480
3'UTR
BASE COUNT 3891 a 3446 c 3680 g 4463 t
ORIGIN
alignment_scores:
Quality: 739.00 Length: 2502
Ratio: 0.645 Gaps: 123
```

```
Percent Similarity: 45.803 Percent Identity: 22.542
alignment_block:
US-09-301-906-15 x BYVUAA ..
Align seg 1/1 to: BYVUAA from: 1 to: 15480
137 ProTyrAlaProValLysArgGluGlnProLysProAlaValLysGlnAs 153
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1062 CCTCGTCTTCCTAAATTCGGGAAGTTGAAATTCCTGCTAGTCAAGAAACA 1111
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
153 pGlu.....GlnLysProLysArgGlnAlaSerHisTrpA 165
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1112 TGTGTGTACCAGCAGTAGTTTCTAAACACAGTTCGA.....A 1146
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
165 laValLysProThrAla...ValGlyValHisValProLeuProLysLys 180
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1147 CGTTTAGACACGCTGCTACCATCGTGGCTGAGTATGTCACACGCTCGGAAC 1196
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
181 GlnGluAlaLeuGluProAlaGlnSerValProGlnGlnSerLeuGluG 197
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1197 CAGTGTTCGCGCGCTCCAAGA.....AATCATCCGATTTTACG 1234
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
197 uLysAlaAlaLeuThrPheGlyLeu.....PhePheS 208
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1235 CAGTGTCTTATATCTTCGTTTAAAGAAATGCGCGTCCAGCGCTTCA 1284
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
208 eLysGlyGlyGlyAspGluSerAspAlaValLeuArgLysGlyLys 224
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1285 TGAAG.....GAAAGAAAGAAATACTACGTGAAGCGCTCTAAG 1322
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
225 LeuPheAsn.....ArgAlaLeuAsnValProIleAsp..... 235
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1323 GTCGTGAGTAGTGTAGTCTAATAAAGTCCTTTGGAAGCTTTGCCAC 1372
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
236 ....ValLysAsnThrPheValTrpAla.....LysI 245
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1373 CATACTAAGAAAGCTTACCACAGTATTCTTACAACTCAGAACGTTAAAT 1422
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
245 leTrpAspGluAlaSerArgArgGlyTyrPheTyrValLysAspArg 261
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1423 TTTAGCAT.....CATTTATCGCGCAGCAGCTTT 1451
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
262 AlaValLysPhePheProIleValArgGlyArgAlaThrIle..... 275
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1452 GAAATTGAAGTCATCCGCTGCGAGCGGCAATTAAGTGTCTTCTTAT 1501
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
276 .....GluAspPheIleValAsnThrAla.....ProGlyC 286
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1502 TTTGCTTAAGGCGAAGCTTACTGCGTAGTCACCTGCGGCACGCGCAGT 1551
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
286 ysAspValAlaLeu.....ProArgIle....Glu 294
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1552 ACCAGCTGCTTTAAGTATGTCGCGTGGGATGCTCTCGCGTGTGTGAA 1601
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
295 LeuTrpSerMetArgGluArgAlaPheValCysThrThrLysGlyTrpCy 311
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1602 CTTTGTGACATACCGA.....CCGCGCGAAGGGTTATG 1633
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
311 sTrpPheAsnAsnGluArgLeuArgGlyGluIleTyrArgArgCysP 328
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1634 TTACCTCGCCCAAGCTGCTCTTTGTTGCGCTCTCAGAAAGCAGCACCTTC 1683
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
328 heSerSerSerPheSerIleGlyPheLeuMetHisLeuGlyPheArgSer 344
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1684 GCGAAGAGACTTCTCGTTGTTGATGTATCCCGACGAAGTTTGTCTTCGCT 1733
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
345 LeuLysValIleArgPheAlaGlyThrAsnIleLeuHisMetProSerLe 361
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1734 AAACGACCTACGTAAAGTGGTCTCTAGTGTCTCAAAACATCCTGTGCG 1783
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
361 uAsnGluGluArgThrPheGlyTrpLysGlyGlyAspVal..... 374
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

1784 CGGAAGACAAAGTCTCTCGTTCACTCTTCCACTGTGATGTGGCTTCGCTT 1833
 375TyrLeuProAsnValProLysThrAlaIleVal 385
 1834 TTTTCGAGCCCTTTTACAGTTTCCCGGTTTC 1865
 386 AlaGlyAspArgThrArgLeuGlyGlyGluI 396
 1866ATCGCGGCGGTGGAAGAACAACTCTGAGAT 1897
 396 eLeuAlaSerValAlaAsnAlaLeuAsnGlnGluValTyrSerSery 413
 1898 CACGTCTCTCTTAAGCACAAGG 1920
 413 alValSerSerIleThrAsnArgLeuValLeuArgAspGlnSerAlaLeu 429
 1921 CGATCGAATCGGTCTACGAACGTGTACGATTCACAAAGACAACTTGTG 1970
 430 LeuSerHisLeuAspThrLysLeuCysAspMetPheSerGlnArgAspAl 446
 1971 GTCGTAGCGTGGAAAAGGATTGTACGACTTTAAAGACGAATCAAGTC 2020
 446 aMetIleArgGluLysProSerHisArgCysAspValPheLeuLysProA 463
 2021 GCTGTCGAAGAAAGGTTCCGTTACCGTCCCTCTATATGGGGAGG 2070
 463 rgGluArgGluLysLeuArgGluLeuPheProGluLeuSerIleGlnPhe 479
 2071 CCGTACAGAGTGTTCAGCGCGCGGTACCCCTCAGTTAATTTGAGTTTC 2120
 480 SerAspSerValArgSerHisPropheAlaAsnAlaMetArgSerCy 496
 2121 ACTCAGAGTGTATTTCGGACCACTTCGCGCGCGCGGTTCCTCGTCTGTT 2170
 496 sPheAsnGlyIlePheSerArgArgCysGlyAsnValCysPhePheAspI 513
 2171 GGAACGAACACTTAGCTCTGATGCTAAATCT...TCCTTTCTGATA 2217
 513 leGlyGlySerPheThrTyrHisValLysAlaGlyHisValAsnCysHis 529
 2218 TAGGGGGCTCCCGCTTTTTCACATAAAGAGGGGACACAGATTACCA 2267
 530 ValCysAsnProValLeuAspValLysAspValLysArgArgIleAsnG 546
 2268 GTGTGAGACCGATTACGACATGAAGGACGCTCAACGAAGAGTTTCGAG 2317
 546 uIleLeuPheLeuSerThrAlaGlyLysSerTyrValSerSerAspL 563
 2318 G.....GAACTCAAGGAGAGGGCTGGTGGAAACCTTCCCGCGAAC 2361
 563 euLeuThrGluAlaAlaSerLysSerValSerTyrCysSerArgGluSer 579
 2362 AACTTGTGAAGCTCAGGCGCG...GTTTCGGTGTGCCCTCACACTC 2408
 580 GlnAsnCysAspSerArgAlaAspAlaGlyPheMetValAspValTyrAs 596
 2409 GGCATTTGAACGTGAAACCGACGTCTAATCATGTGTCAGGTTTACGA 2458
 596 pIleSerProGlnGlnValAlaGluAlaMetAspLysLysGlyAlaLeuV 613
 2459 CGCATCTTTGAACGAATTCGTCGGGATGGTCCCTAAAGAAATCGAAG 2508
 613 alPheAspIleAlaLeuMetPheProValGluLeuLeuTyrGlyAsnGly 629
 2509 TCGCTTACCTCACCATGGTCACTCCGGGTCAACTTCTAGATGAACGGAA 2558
 630 GluValTyrLeuGluLeu.....AspThrLeuValLysArgGlu 644
 2559 GCCTTCGTATCCACCTTTGGGGTGGCGACGTGCTGTCGGACACCCGCG 2608
 644 yAspTyrLeuAlaTyrAsnValGlnCysGlyGluMetTyrGluHis 661
 2609 AGACATGGTGCATACAAATTCGA.....TCATCTTGTACTGCCACA 2652

661 erPheSerAsnValSerGlyPhe.....PheThrPheSer 672
 2653 AACTGTCCAATATAAAGAGTATAATGTGACTCCCGCATTCACCTTC... 2699
 673 TyrValArgThrSerSerGlyAsnValPheLysLeuGluTyrGluGlyTy 689
 2700ACGGGAATCTTTTTCGGTCGAGATGTACGAAA 2734
 689 rArgCysGlyTyrHisHisLeuThrMetCysArgAlaGlnLysSerProG 706
 2735 CCGAATGGCGTCAACTACTACAAATCACTCGTTCAGCGTATTCCTCG 2784
 706 ly.....ThrGluVal 709
 2785 AGATACCGCGGTTAACACGCTGCGGTACCGCCGAGGCTGCACCGAGTC 2834
 710 ThrTyrArgSerLeuValProSerPheValGlyLysSerLeuValPheI 726
 2835 GTGCAAGTTAACTG...CCACGCTTCGATAAACTCTAAAGACGTTTCT 2881
 726 eProValValAlaGlySerSerValSerPheLysThrIleValLeuAsps 743
 2882 CTCC.....GGGTATGATTACATATATTGGATG 2910
 743 erAspPheValAspArgIleTyrSerTyrAlaLeuAsnThrIleGlyThr 759
 2911 CGAAATTCGTCTCGAGTCTTCGATTACGTGGTTCAGCAACTGTAGCGTT 2960
 760 PheGluAsnArgThrPheGluTyrAlaValGlyAlaValArgSerGlnLy 776
 2961 GTTAAACAGTAAACCTTCGAGTGGGTGGAGTTACATAAATAACAGTAA 3010
 776 sThrHisValIleThrGlySerArgValValHisSerLysLysValAspI 793
 3011 ATCGCGGTTGTCTATAAGTGGAAAGTGATACATCGAGATGTGATATCG 3060
 793 erProAspAspMetTrpGlyLeuValValAlaValMetAlaGlnAlaIle 809
 3061 ACCTCAACACCTCTGAATGTTTTCGGCTGTGTGCTCGCGTGGGTG 3110
 810 LysAspArg.....AlaLysSerIleArgSerTyrAs 820
 3111 CGTTCTCCACAACTACTGAATTTTACGGAAGAACTCAATATTATACAC 3160
 820 nPheIleLysAlaSerGluGlySerLeuAlaGlyValPheLysLeuPhe. 836
 3161 TGGCGATGCTTCGTGCTTTGAGACTATCCGTTTCTCTTCGGGAGTGA 3210
 837PheGlnThrValGlyAspCysPheSerAsnAlaValSer 849
 3211 GCGGAGGCTTACGCGGAATAAATCCAGTTCGAAAGCTTATGAAG 3260
 850 ValTyrAlaLysAlaMetValHisAspAsnPhe.....AsnVa 862
 3261 AGCATCTTTCCGCTGGTGTAGATTACGAATTTCTCGATCTTTGACAATC 3310
 862 lLeuGluThrLeuMetSerMetProArgAlaPheIleArgLysValProG 879
 3311 GCTCCACACTTGTCTGAATACCTCAGAGTTGAAGTGGCG... 3350
 879 lySerValValValThrIleCysThrSerGlyAlaSerArgLysLeuGlu 895
 3351GTTTCAATCGCTCAGAATGCTGAGGTGGAT..... 3380
 896 LeuArgGlyAlaPheAspIleSerLysGluThrPheGlyArgLysLeuLy 912
 3381TG 3382
 912 sAsnSerArgLeuArgValPheSerArgAlaIleValGluAspSerIleL 929
 3383 CACGAAGAACCGTGTGTTGACGGAGATAATAGCCGAA..... 3422

5094 TTCTTTAAACCTCGCTCATACCGCTTCGGAATGTGTGTCGCAGATATC 5143
1371 GlyVala 1373
5144 CTGCCTGCCTCTATTTCGCTTATGTGGGGACACACGAGCTGTGTTAG 5193
1373 spAspValPheHis.....GlnGlnSerSerGlyGluThrAlaArgGlu 1387
5194 TTAGGTATTTTACATAACTTATTCAGTACTCGCGTGTGTCGCTGTT 5243
1388 Val.....GluValAspG 1392
5244 GTGTCGCGGTGCTGTATGCCCTCTTTTGGTCAGAAACAACTGCATG 5293
1392 Y.....LysGlyLeu..... 1395
5294 CTGAGTATGATTAACCAACATGGCGCTTTCGGTTTTTCGTCGCTTGT 5343
1396LeuProGluSerValValGlyGluAlaProThrGlnGluArg 1409
5344 TCGTCGCCCTTCGGCGCGCTTTCGGCGTACTCTAATTCGGCCCTACGA 5393
1410 GlyArgAlaAlaAsp.....GlyAsn.....Th 1417
5394 CGGAAGATTATCGAATTCAATTTTCGGAACATTCACCACCTTTTCATCT 5443
1417 rAlaGlnThrAlaValAsnGluGlyAspArgGluProValGlnSerSerL 1434
5444 CGCAGTAATCGAAACAAAGAGTTTCGCGCGGAACCTCTT..... 5483
1434 euValSerSerProGlnAlaAspLeProLys.....ValThr 1446
5484TCACCGGAAGTAGACATTTGATGTCGACTGCGACTTGTGTTCC 5525
1447 GlnSerGluValHisAlaGlnLysGlnGluValLysGlnGluValProLeuAl 1463
5526 GATTCAGATCGGTTCTTCAGATGAAGTCGCGTCGAAT...CCCGTCC 5572
1463 aThrValSerGlyAlaThrProIleValAspGluLysProAlaProServ 1480
5573 AGGCTTTACACGGTGGGAGT.....AGCGCTCTTCCAACT 5607
1480 aThrThrArgGlyValLysIleAspLys.....GlyLysAlaValAla 1495
5608 TTCTAACCTCTCTCGAAGGTGTTTTAAATTTGCTAGGCGCATTCGCG 5657
1496 HisValAlaGluLysGlnValGlnValGluGlnProLysGlnArgSe 1512
5658 CGTTTGTCTTTCGTTTGGCAATTTTGGCGCTACTTTGTTCAACGACG 5707
1512 rLeuThrIleAsnGluGlyLysAla.....GlyLysGlnLeuC 1525
5708 GCTGGCTTCAAAAAGGTTGAAGACATTCATCGGCTCTGGCAGGCTGTTG 5757
1525 yMetPheargThrCysSerCysGlyValGlnLeuAspValThrAsnGlu 1541
5758 ATAATTCTCACTCACTCGGTGGTTTACTCTTCCAAAGAGTACGATTCA 5807
1542 AlaThrIleAlaThrArgPheSerAsnAlaPheThrPheValAspAsnLe 1558
5808 GTGCTG.....AACGCATTTATAGACGTCGAGCTAAT 5839
1558 u.....LysGly..ArgSerAlaValPheSerLysLeuGlyG 1571
5840 CCTACTTAACCTCGGTAGCGTAATGCTACTCTTACTTCTTCTGGTCA 5889
1571 uGlyThrThrTyrAsn...GlyGlySerHis..ValSerSerGlyTppPro 1586
5890 GGGGATCTCTAACGAAGTTAGCGGAAGCATCGTTGGTTTCAGGCTTCGCT 5939
1587 ArgAlaLeuGluAspIleLeuThrAlaIleLysTyrProSerValPheAs 1603
5940 TCGTTTCTAGGAAGGATGTGC..... 5960

1603 pHisCysLeuValGlnLysTyrLysMetGlyGlyGlyValProPheHisA 1620
5961TCTCGTGTCTGCTGCTG... 5978
1620 laAspAspGluGluCysTyrProSerAspAsnProIleLeuThrValAsn 1636
5979TGCTCCTCTCTTCGAAC..... 5996
1637 LeuValGlyLysAlaAsnPheSerThrLysCysArgLysGlyGlyLysVa 1653
5997GCCGTTGTAACTTATGAGTCCGCTTCGTACGAAAGGAAG... 6038
1653 lMetValIleAsnValAlaSerGlyAspTyrPheLeuMetProCysGlyP 1670
6039 .ATCGATTCCCTCTCTCTCCGTTTCAACCGCTTCGATG.....T 6078
1670 heGlnArgThrHisLeuHisSerValAsnSerIleAspGluGlyArgIle 1686
6079 ATGAACGT.....CTCGAAGCTCTCGAGAGCATATCCCGCAACACGTG 6122
1687 SerLeuThrPheArgAlaThrArgValPheGlyValGlyArgMetLe 1703
6123 CTTTCCACGCTGCTGT.....GTAGGA..... 6143
1703 uGlnLeuAlaGlyGlyValSerAspGluLys..... 1713
6144AGCGACGAAGAAGAAAGGCCGAAAGAG 6174
1714SerProGlyValProAsnGlnGlnProGlnSerGlnGlyAlaThr 1728
6175 TGACAGAACCGAATTTGAACATACTCTCGAAGATGTTGTTCTATTCGT 6224
1729 ArgThrIleThrProLysSerGlyGly..... 1737
6225 TCACACTCGCAACCTTATCTGGAGGTGAATGTTCTATTCTCAAGATCG 6274
1738LysAlaLeuS 1741
6275 TGAAGAGAATGAACGACGCAACCTGTTACGCGACGTTAGCAAAATCGTCA 6324
1741 erGluGlySerGlyArgGluValLysGlyArgSer...ThrTyrSerIleT 1757
6325 GTGAACGAAGGGTTTGGAGACCGCGCGTGAACAAACACTACTCTACAT 6374
1757 rpCys..GluGlnAspTyrValArgLysCysGluTrpLeuArgAlaAspAs 1773
6375 GGTGTAAAGC.....GAGTTTCTCAACGCTATTAA 6403
1773 n.....ProValMetAlaLeuGluProAspTyrT 1783
6404 TACTAGCAATGAGCAACCTAGGCGGATAATCGTT.....GACCAC 6444
1783 hrProMetThrPheGluValValLysThrClyThrSerGluAspAlaVal 1799
6445 CTCGCCGAATCT.....CGCGGTTGACCAACTCCGTCGAGGAA 6482
1800 ValGluTyrLeuLysTyrLeuAlaIle.....GlyIleG 1811
6483 TTTCTATTACTCCAGGAACCTCGCTCTTTTGAGTTGAGTTGCAAACTTCG 6532
1811 uArgThrTyrArgAlaLeuLeuMetAlaArg..... 1821
6533 TGAGTATTACGATCAATTGAAGGTTGCGAATTTTAACAGACAAGAGTGT 6582
1822AsnIleAlaValThrThrAlaGluGlyVal 1831
6583 TGTGCGCAAAAGACGAAGACATCTTTGTTCTACGAGCGCGCAAGGTGA 6632
1832 LeuLysValProAsnGlnValTyrGluSerLeuProGlyPheHisValTy 1848
6633 GTTTCGCGGCAAGAAC.....TCGAGGTTGCCCTCTTAAGCATTTC... 6671

```
1848 rLysSerGlyThrAspLeuIlePheHisSerThrGlnAspGlyLeuArgV 1865
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
6672 ...AAGGTCACGAATTTTGTTCGATCT.....GGAGGGTTG.... 6707
1865 aIArgAspLeuProTyrValLeuIleAlaGluLysGlyIlePheThrLys 1881
      ::::: :::::
6708 .....GTCCTTAC..... 6716
1882 GlyLysAspValAspAlaValIAlaLeuGlyAspAsnLeuPheValCy 1898
      :::::
6717 .....GACGGTACCAGCAGACT 6733
1898 sAspAspIleLeuValPheHisAspAlaIleAsnLeuIleGlyAlaLeuL 1915
      :::::
6734 GGACACCAATT.....TTTCACACCAAAACGAATTTCCGTCGCG...A 6774
1915 ysValAlaArgCysGlyMetValGlyGluSerPheLysSerPheGluTyr 1931
      :::::
6775 AGCGCGCTTCTTCGGGCTATCTC.....TCTATAGAACTTTTCACCTTC 6818
1932 Lys.....CysTyrAsnAlaProProGlyGlyG1 1941
      :::::
6819 ACTAATTTTCAGCGCTAACGTACTGCTGACGAAGCTCTCCAGTGGTGG 6868
1941 yLysThrThrThrLeuValAspGluPheValLys.....SerProAsnS 1956
      :::::
6869 AAAGACCAACGACTTTGATAAAGGTTTGTGCAAACTTTTCAAAAGTTA 6918
1956 orThrAlaThrIleThrAlaAsnValGlySerSerGluAspIle..... 1970
      :::::
6919 ATTCTGCTGATTTTACGGCGAATAAGATTCCGGAGAGAATAACTTCGG 6968
1971 .....AsnMetAlaValLysLysArgAspProAsnLeuGluGlyLe 1984
      :::::
6969 AAGGTGAATCGCATCGTACTTCACGAAGCGATACCCCTCTTCACACCGG 7018
1984 uAsnSerAlaThrThrValAsnSerArgValValAsnPheIleValArgG 2001
      :::::
7019 TCACAGGATTTTAACATCGATTCTATCTAATGAAC.....AACAGAG 7062
2001 yMetTyrLysArgValLeu...ValAspGluValHisMetMetHisGln 2016
      :::::
7063 GTTTCAGCTGTAAAGTTTGTACCTCGATGAGTGTTCATGGTTCACGCT 7112
2017 GlyLeuLeuGlnLeuGlyValPheAlaThrGlyAlaSerGluGlyLeuPh 2033
      :::::
7113 CGAGCTGCTGTAGCTGTGATCGAATTCACCAATCGGATTCAGCATCTT 7162
2033 ePheGlyAspIleAsnGlnIleProPheIleAsnArgGluLysValPheA 2050
      :::::
7163 ATTTGGACAGCAGCAGCAAAATTCGTTAC.....GGTCGATGACTG 7203
2050 rgMetAspCysAlaVal.....PheValProLysLys 2060
      :::::
7204 AATTGGATGACTGCTGCTTCTTCAGATTTGAACCGGTTTGTCT....GATGAC 7250
2061 GluSerValValTyrThrSerLysSerTyrArgCysProLeuAspValCy 2077
      :::::
7251 GAATCGACAGTATATGGTGAACCTCTCATACAGGTCTCCTTCGGAGCTTG 7300
2077 sTyrLeuLeuSerSerMetThrValArgGlyThrGluLysCysTyrProG 2094
      :::::
7301 TCTTTGGTTGTCAACTTTC.....TACCCGA 7326
2094 LuLysValVal.....SerGlyLysAspLysProVal 2104
      :::::
7327 AAAGTGTGGCCACTACCAACTGTTGTTTCAGCCGGTCAATCT..... 7367
2105 ValArgSerLeuSerLysArgProIleGlyThrThrAspAspValAlaG1 2121
      :::::
7368 .....TCGATCCAAGTACGCGAGATTCAACCGTAGACGCTCGAATA 7411
2121 uIleAsnAlaAspValTyrLeuCysMetThrGlnLeuGluLysSerAspM 2138
```

```
      :::::
7412 TTCCAGTGAATTCGCTACTTCTACTATGTTTACAGTCAGAGAAAGATC 7461
2138 et.....LysArgSerLeuLysGlyLysGlyLysGluThr 2149
7462 TCCTGAATCTTTTCGCAAGAGGTCGCTTCGAGCGTTGAGAAACCCAGC 7511
2150 ProValMetThrValHisGluAlaGlnGlyLysThrPheSerAspValVa 2166
      :::::
7512 ...GTCTTAACAGTCCATGAAGCTCAGGCTGAACCTACCGCAAGTTAA 7558
2166 lIleuPheArgThrLysLysAlaAspAspSerLeuPheThrLysGlnProH 2183
      :::::
7559 CCTCGTCAGACGAAATTTCAAGAGGACGATCCTTTTCGGAGCGAACC 7608
2183 isIleLeuValGlyLeuSerArgHisThrArgSerLeuValTyrAlaAla 2199
      :::::
7609 ACATCAGTGGCCTTGTCTAGCATGTGCGAGAGTCTGACCTATTTCGGTC 7658
2200 LeuSerSerLeuAspAspLysValGlyThrTyrIleSerAspAla 2215
      :::::
7659 CTGAGTAGTAACGTGACGACCAATAGCCCAAGCTATAGTGAAGCG 7706

seq_name: gb_vil:CTU16304

seq_documentation_block:
LOCUS CTU16304 19296 bp RNA VRL 15-MAR-1999
DEFINITION Citrus tristeza virus complete genome.
ACCESSION U16304 U02547 L20760
VERSION U16304.1 GI:806738
KEYWORDS
SOURCE Citrus tristeza virus.
ORGANISM Citrus tristeza virus
VIRUSES; ssRNA positive-strand viruses, no DNA stage:
Closteroviridae; Closterovirus.
REFERENCE 1 (bases 12006 to 19296)
AUTHORS Pappu,H.R., Karasev,A.V., Anderson,E.J., Pappu,S.S., Hilf,M.E.,
Febres,V., Eckloff,R.M.G., McCaffery,M., Boyko,V., Gowda,S.,
Dolja,V.V. and Koonin,E.V.
Nucleotide sequence and organization of eight 3' open reading
frames of the citrus tristeza closterovirus genome
Virology 199 (1), 35-46 (1994)
REFERENCE 2 (bases 10009 to 13033)
AUTHORS Karasev,A.V., Nikolaeva,O.V., Koonin,E.V., Gumpf,D.J. and
Garnsey,S.M.
Screening of the closterovirus genome by degenerate primer-mediated
polymerase chain reaction
J. Gen. Virol. 75 (Pt 6), 1415-1422 (1994)
REFERENCE 3 (bases 1 to 19296)
AUTHORS Karasev,A.V., Boyko,V.P., Gowda,S., Nikolaeva,O.V., Hilf,M.E.,
Koonin,E.V., Niblett,C.L., Cline,K., Gumpf,D.J., Lee,R.F.,
Garnsey,S.M., Lewandowski,D.J. and Dawson,W.O.
Complete sequence of the citrus tristeza virus RNA genome
Virology 208 (2), 511-520 (1995)
REFERENCE 4 (bases 1 to 19296)
AUTHORS Karasev,A.V., Hilf,M.E., Garnsey,S.M. and Dawson,W.O.
Transcriptional strategy of closteroviruses: mapping the 5' termini
of the citrus tristeza virus subgenomic RNAs
J. Virol. 71 (8), 6233-6236 (1997)
REFERENCE 5 (bases 1 to 19296)
AUTHORS Karasev,A.V.
Direct Submission
Submitted (24-OCT-1994) Alexander Karasev, University of Florida,
Citrus Research and Education Center, 700 Experiment Station Road,
Lake Alfred, FL 33850-2099, USA
On May 16, 1995 this sequence version replaced gi:456172 gi:508435.
FEATURES
location/Qualifiers
1..19296
/organism="Citrus tristeza virus"
```


/strain="T36, quick decline"
/specific_host="Citrus sp."
/db_xref="taxon:12162"
1. .107
5' UTR
108. .9482
CDS
/note="5' UTR of CTV has approximately the same size as the respective UTRs of other closteroviruses, however it has no sequence similarity to them"
/codon_start=1
/product="349-kDa viral polypeptide"
/protein_id="AAC59623.1"
/db_xref="GI:806739"
/translation="MSKIRGFWSSAIAVNSDYTIISRIRKRLNTIVTVLHYGFCVRIK
VIRKSDADPTPLKRSVPFESVRAATSVLLGLVLLRLTQGVAFPLTSLHALFV
PALPAGFSRLCKPPLTSAKRSLROAKRESVLSRRASRFSREVRGTRSLRHR
VVRADTAPVSPQDKPTRNSRLGKTHLPYVGEAALDDILRWEEVNPSPVAIP
VPTFTGTPATAMCSAALRAALTITRLKLSPLRRNTYGTISIMTYLLPLTPVPR
FLPASATFSPSKLGLVEVPSGRDLHTAFVSSHASAFFSDAVRASFPVSAEQPFH
RQKTHVLAITDPLCGTVSLMSDGVNTARLDVVAIYAYAESLGARVRFWMESD
VDDAKRVQCYVHRFDVALYFGRRADLVRRALGMPTVYALKALVYREVGRSLK
VPMRGTTFGSGVHCLSLNLOVRLSIPNHLVGTSTSAHLVIGOVRSVQGISVPSR
SAVRNRNRDGFGLSGSSSSNGESTRCSDSSPGLSPSPKLSGLRISIDR
RFRPAGTQLRVVRVDDHYVSTGNSDELCHAVLRMSKGLTLLILHSIRERK
GLLGAACFRGSSSHVFKGNFTAREPISAVRYENYAAVMDDVGTWPKGYVPDY
DVGKRRRSRNFPLSHRAVDSPALQSSPVEVASVSDKPSKOKASPSSTSSSS
FNSVRSDDWNVGSHRTNKRPREYAGHRRIGNFTPRGTVYNTVDERAYKRLR
LDRDTACSLILLARLSGYPLSDDEFKCVTSFVACIEPVESGLVHFERSDVR
ASFPDGOPIHATALTILEVVVSDMWYAFQNPCLKDGYCYIRHFAVSLMSGRIFR
RDVLDGPPVFEVOHRLERLYKAALRYGVQTSAPRCFHCYNDSPRPMASFN
KMGDESLNLTIDRLRAGVSAIEVKVRIPOSLVRSVEKTLIEQNTLVDMHRS
RYSVVPVSENOQVMTLTPYEFNFIHVSVDHPVAAGSRALENHLVRKKGATD
YSDVGGPFLHLAGHGVHPCVPYVDKDAHRVVRHQLSKVSDQSDGVKVGWT
VNTSVCCNIGLCEYHASEAMVQVDPVPLRELCRAMINKKTSVCYMTWTPGELL
ARESEFIDKDCVELDPIADRVYCFNNSAYTHYSTICECMRTPCLVLDGFLPTIE
MVSRLCSVNYCYTKSSVCPRISETKRLRYRCDSDLRIRIKPRYSKTRSCLP
LYDAKFSVRVYVYVNNVNNVNTAFEMTWNYSKSKSVISGKIIRDEPTALEY
LDGFSVAGVAGQNVADYVNRRAAFSGETSIFELVGAISEKCRDLFLGIQDAVE
RCVKSYLEFNSMFLSDLLTISEYSELDVLDLPFGGIVDGMFVRLMEDGVKS
ALLKAVOAEALSKRGNTSEIIPVCDSDSDNGDDSDRVSDTDMFVRLSGSGLR
AGAPSLFRCILLKGLGGLCAKLVLELLVYRVESFSLRKRKTLARLTSFANLLARL
SCGFDDFTVFAWELIAINLRIVDGVVVKLGCPRNLFRLIPKLACVAKDVKDVS
AYRVSFAVLGKSKYKLCQPCLEIFDGLIEVSARNLGLTCVSTALAI FDLVIT
PLGASVPPVCAVFRGLAMEVSNFLFKVFGKPTLGEDTFRVSMVLRKAFD
KLSVDAGLVCVGPVIRIILSLFSEDDAVGYSKHEVSLPIVTFVCKVRLSY
LERCKEACVAFVDSAKILKSLTSEFNESIONSDLTSGAKSLGHVSSILYSSGVI
NKASAVATVAVGSDISGRVKNRPFSLSRHDEQTSYISASDCSELDESFS
PGLRGNACRDLFRIVSRRAVACCEYVTLILKRIASRAGEYSYVMTPLVRLVYATD
DVRRCSRFTDCTDKASNFVTGLSMRSLRSGAYSATPGLGGSGRGLTSFHTGGFT
RVIGALNTLSVDFDEEFRTLHLRADVVLLVEFLKLPVCLFSTYITIECRGRI
PLREFLWFCRFVMDYTEGLSINTYOLSAVEKYVYRGNRENEDETLRVITE
AEGRIEFLVLAEMERPAASNEVSSSSSDSPFPYTGIELELNDSDSALVIG
VDNGTGRITVGETDSAGSSSDVSLPDRDFEQIRGLGNGVDVRLGLKFLVLLK
NLFDARSFVLYLGLLTLYFGGYNKFCVRLFSLGTCPLFEFLRACFASARLLK
LASVYQTNLSVEYVHLFVSRLLYDYSRYFATVPSLERLNFYLDYFLSLPRLYF
PNRSARSCARPVPVIRTEGVAPRERSLNDLSDLSKVSOLKGGTFFVDDGESD
GSEHSIGSIVGNEVNLNAGSDVCEKAPLMKHVTALSLRHPHSRGRDIPVSRK
TFVRESSEGVSKVSMCDYLTLASSECTVPMYVARDITYKLTINAMREFFYSQKI
TLYELHGKLSYBDELKAVGDRKLTMDLDDVTVDFNRLKLVGRNGSHPRNTLK
SYCMFCNDGLVPCRSQKYALISNQTSEFVANSFLRAVDGKDLTFTNEHSLYV
EAPCGGTHSLVNSADYCVKSLVVTANKNSQTEISQISNLMGRKILAAKYVTD
AASRFTVDSVLMNHLRTQLLFTDECFVHAGAIGAVEFTSCVAKVFFGDSKVOIH
YIHRDLGVSAVDADAFIOPEHRIYGEVSYRCPWDICEMLSEYPRHIVATANYGISG
KSSVIEINGCDVDPYDKAKYIVYTAQAKNDLQKGLRLTVGRKNVPIVNTVHEV
QGEYKRVLRVRFQEDTFFSSKNHIVVALTIRHVDLSVYSLTSRRYDDPATNIDRA
KEIFDKPSSNHSFGSSLEWYLEKYPTEYKGSKASSAPISCIENEFNEVVGSSVQ
LGDVSEELSSRPFSGCDNVTRDTAPDPSGNLHPEPARVRRSKVTSSSKSKSVARKP
SVWRKS"
-9355. .10857
/function="replication"

CDS

/note="RdRp: RNA polymerase homologous to RNA polymerases of other closteroviruses, it is presumably expressed via a +1 ribosomal frameshift"
/codon_start=1
/citation=[2]
/product="57-kDa RNA-dependent RNA polymerase"
/protein_id="AAC59624.1"
/db_xref="GI:3337235"
/translation="ETPPLLTRVYTNRAFCGVVRSSQAIPPRKASLQENLLSYESRNY
NFKTERVGPSEFGRAAAVIERCFKMEEMAKIRCDIISLTEAMILKWLDRTPCQ
IKAVHGEKLPSEVQEQSNFKLMVKROAKVLDLSSSKHHPAAQINMFKHFIINAIF
SPCDEKFNRLVSLNDINIVFTMTNAGLAPIIRIIGDDNDLFGVEYDFSKFKQSO
DLFKEERTLYSEGFDELLDVMGEYRATATLDQLSFSVDGQRSGSGNTWI
GNSVLTLGSUTYDYSKFDLGLVSDGDDSLYSSKISNFSSEICLETGFETKFMSPS
VYFCSEVWQTKNTCPVDPYKLLVKGAPONKLTDFELFELTFSKDMTQDFDQ
VLEKLLKLCERVAYFASQTPALCAIHCVRSNFLSFLRFPFTRGWTVDALALRQ
LRKLTNLICERVYVYDNRVSYFSDNPTFKPDANDNDVLDLQAGELATG"
10906. .11814
CDS
/note="p33: The 33-kDa protein seems to have no homologs in genomes of other closteroviruses"
/codon_start=1
/citation=[2]
/product="33-kDa protein"
/protein_id="AAC59625.1"
/db_xref="GI:806741"
/translation="MFAFASQDLLEKVFRRRTYHRYKFGDVVKDFTIDIGYDFTD
RDPYAGDYLSLYFLFNNDVSGPIAASIVSPVSPVSGTHKIRAHIDNQPCNGVTV
RTDKSRVIRVYKASPMRGYYSFRVFLSSDVASRSEFVSCFVFLCCCTQYRST
NLSKVCSSFPFRVASETATNEFVATDDVEDVYIRKQAGEFSCADYPPRCYRST
NLGDSGVQSRITIEEGYMTGAGNAVTVPTPLVASVPYIAGYNENARSRVSLIR
RVGCYAVCLVSVLMSGLLAIIF"
11880. .12043
CDS
/note="p6: A small hydrophobic protein, probably membrane-associated, homologous to small hydrophobic proteins of other closteroviruses"
/codon_start=1
/citation=[2]
/product="6-kDa protein"
/protein_id="AAC59626.1"
/db_xref="GI:806742"
/translation="MDCVQGLTFLVGIACAFAGLIIIVITVRCYIKPVRASAP
YGTATHV"
12049. .13833
CDS
/note="p65: CTV-encoded homolog of HSP70 family of heat shock proteins, identified by sequence comparisons"
/codon_start=1
/citation=[1]
/product="65-kDa protein"
/protein_id="AAC59627.1"
/db_xref="GI:806743"
/translation="MVLGLDFTTFTVAMATPSELVTLKQSSSYIPTCLLHAEP
NSVSGYDAEYLAASGESGFYKDLKRWGCTAKNYQYVHLKLSYKVIKVEGCTKS
VPVPLSLNLDGLSLVALPSLIASAKSILSDAERFNVSTGVICSPAGYNLQR
ATQOSISMSGYSCVYIINEPSAAATSTPLKUNLSADKLAIVDFGGTDFSDIVSRL
PTFAVSSGDMNLGGRDIDKDKIYEMADFPQKELNWSLSEALSLQDPVRYT
VNHGMSTVSDITVRLIASVFNRTIDILTVKSSMPESQSLKLVVYGGSYL
PCLDALATVPVSGIVPEVDEATATVATCGLSCLDRSKALLIDCTHHLVSTTF
SADSVVAAAGSIPFEGEOKLTLRKCVSTSNYQARMEFGDYKVFERNERYAASISL
FTLGVNWSVNDVENTLTKVDSMGKVEYFLKPSGELVNVQGTSHDYAGHPHTRK
LVRLSDYNVNSALVALTLTREKREKFLRLFDTLADLRKLTASLGEYSKKYPIIR
NIDVVSRRMGIVSVKLRGSDLEIRPL"
13757. .15364
CDS
/note="p61: Hypothetical protein gene with unknown function has similarly positioned homologous genes in other closteroviruses"
/codon_start=1
/citation=[1]
/product="61-kDa protein"
/protein_id="AAC59628.1"
/db_xref="GI:806744"
/translation="MSSHVWGSFLRFKYGEAIWKEYLSESTRNFDERNVSLDHTLSS
GVVVRRLNAPQGTTFENALLYNVVDVDFELTGMLPLKMTGTEDRKPVELI


```

465 ArgGluLysLeuArgGluLeuPheProGluLeuSerIleGlnPheSerAs 481
    |||
3240 CAGGTATGCTTACTAGCGGTACCGGAATTAATTAATTAATTAATCA 3289
    |||
481 pSerValArgSerHisProPheAlaAsnAlaMetArgSerCysPheA 498
    |||
3290 TTCGGTACATTCGGATCCGGTGGCAGCAGGTTCAGGGCGTGGAGA 3339
    |||
498 snGlyIlePheSerArgArgCysGlyAsnValCysPhePheAspIleGly 514
    |||
3340 ATCATCTCGTTCGTAACACACGCTGGT...ACCGATTATTTCGGATGCGGG 3386
    |||
515 GlySerPheThrThrHisValLysAlaGlyHisValAsnCysHisValCy 531
    |||
3387 GGTTGCGGTGTTTATTACGTGGGGTCACTCCGGGTTCCACGTTTG 3436
    |||
531 sAsnProValLeuAspValLysAspValLysArgGlieAsnGluIleL 548
    |||
3437 TAGACCGGTGTACGATGTTAAAGACGCTCATCGTAGTAGTCCGTCATC 3486
    |||
548 euPheLeuSerThrAlaGlyGlyAspSerTyrValSerSerAspLeuLeu 564
    |||
3487 ACCAACTTAGTAGGTTTGGCTAGAT...CAGAGTACGGGTGTT 3527
    |||
565 ThrGluAla...AlaSerLysSerValSerTyrCysSerArgGluSe 579
    |||
3528 AACAAGTTGGTGGACCGGTAAACAGAAATTCGTATGGCGTAACATTT 3577
    |||
579 rGlnAsnCysAspSerArgAlaAspAlaGlyPheMetValAspValTyTA 596
    |||
3578 GGGAGAGTGTATCATGTCATCGAGCGGATGGTTATGGTTCAAGTTTATG 3627
    |||
596 spIleSerProGlnGlnValAlaGluAlaMetAspLysLysGlyAlaLeu 612
    |||
3628 ACGTACCGGTACGTGAGTGTGTAGGCTATGATAACAAGAAACTTCC 3677
    |||
613 ValPheAspIleAlaLeuMetPheProValGluLeuLeuTyrGlyAsnG 629
    |||
3678 GTGTGTATATGACCAAGTTACACGTGGTGGTGTGGACCGCCGCA 3727
    |||
629 yGluValTyrLeuGluLeuAspThrLeuValLysArgGluGlyAsp 646
    |||
3728 GTCCTTTTTATTAAGGACTTGATGTTCCGTT...GAGTTAGATC 3771
    |||
646 yrLeuAlaTyrAsnValGlyGlnCys...GlyGluMetTyrGluHis 660
    |||
3772 CCATTGCAGACAGGGTGTGTACTGTTTCAACAATTCGCTGATACACAC 3821
    |||
661 SerPheSerAsnValSerGlyPheThrPheSerTyrValArgThrSe 677
    |||
3822 ACGTACTCCACTATATGT...GAGTGTATGCGAACTCC 3856
    |||
677 r...SerGlyAsnValPheLysLeuGluTyrGluGlyTyTA 690
    |||
3857 TTCCTTGGTAGTACGAGGGTTTGTGTTTACTATCGAAATGGTATCGCTCA 3906
    |||
690 rgCysGlyTyrHisHisLeuThrMetCysArgAlaGlnLysSerProGly 706
    |||
3907 GGTGTACGGTAAATTAATTATTATGATCACCAGTCGTCGCTCTCTCCGAGA 3956
    |||
707 ThrGluValThr...Ty 711
    |||
3957 ATTATGTCAGACTAAGCGGTGCGCTACCGTCCGTTGTCATAGCGACTTGAT 4006
    |||
711 rArgSerLeuValProSerPheValGlyLysSerLeuValPheIleProV 728
    |||
4007 TAGAATCAAAATTCGCCGGTACTCCAGTAGACTCGTAGTTTACCT... 4055
    |||
728 alValAlaGlySerSerValPheLysThrIleValLeuAspSerAsp 744
    |||
4056 ...GGGTGTTACTATTGTTACTTGGAGCCCAAG 4085
    |||
745 PheValAspArgIleTyrSerTyrAlaLeuAsnThrIleGlyThrPheG 761

```

```

4086 TTTGTAGTCGCGGTACGAGTACGTGGTCAATAACTGCCTTGTGGTGAA 4135
    |||
761 uAsnArgThrPheGluTyrAlaValGlyAlaValArgSerGlnLysThrH 778
    |||
4136 CGCGAAACTTTTGTAGTGGACTTGAATTAATCAATCTTCAAGTCGA 4185
    |||
778 isValIleThrGlySerArgValValHisSerLysValAspIleSerPro 794
    |||
4186 GAGTGTGTATCAGCGGAAGATTATACACCGAGACCAACCAATCGCTTTA 4235
    |||
795 AspAspMetTrpGlyLeuValValAlaValMetAlaGlnAlaIleLysAs 811
    |||
4236 GAGTACTTAGATGGGTTTCTCGGTATGCTGTCCGCTGGCGTGAAGG 4285
    |||
811 pArgAlaLysSerIleArgSerTyrAsnPheIleLysAlaSerGluGlyS 828
    |||
4286 TAGCGAAACGCGACGCTTCTAGACGC...TTAGCGGCTTTTTCGGGTG 4332
    |||
828 erLeuAlaGlyValPheLysLeuPhePheGlnThrValGly... 841
    |||
4333 AAATTCG...ATTTCGAGTTAGTGGGTTTTCGTATA 4367
    |||
842 ...AspCysPhe...SerAsnAlaValSerVa 850
    |||
4368 TCCGAGAAGTGCAGCTTGTCTTGGGTATTTCAGGACGCTGTGGAAG 4417
    |||
850 lTyrAlaLysAlaMetValHisAspAsnPheAsnVal... 862
    |||
4418 ATCGGTGAAGTCGTACCTCAGGAATTCGTTTAAATATGAGTTTCGTGACT 4467
    |||
863 ...LeuGluThrLeu 866
    |||
4468 TGTCGACCCCTTACTTACGATTAGCGAATATTTCAGAAATTCGACGTTCC 4517
    |||
867 MetSerMetPro... 870
    |||
4518 ATCGACTTACCGGGTTTCGGTGGAAATAGTCGACGGTAACGAAACTAGACT 4567
    |||
871 ...ArgAlaPheIleArgLysVal... 877
    |||
4568 TATGGAAGACGGTGTAAATCGCCTTATTACGTAAAGCGGTTCAACCG 4617
    |||
878 ...ProGlySerValVal... 882
    |||
4618 AGGCTCAGTTCAGTCTTAAGAGAGGCAATACTCCGAAATTCGTTCCGTGT 4667
    |||
883 ...ValThrIleCysTh 887
    |||
4668 GATGACGATGACGACTCGAAGCGGTGATGATTCGCGTGTTCGCTCGATAC 4717
    |||
887 r...SerG 889
    |||
4718 GGACTGGTGGTTCCTTCCTCGGGTGGTAGTTCCTGGCGGGTTCGGAGCTG 4767
    |||
889 lyAlaSerAspArgLeu...GluLeuArgGlyAla 899
    |||
4768 GAGCCGCCCTTCTTCTTGTAGTGTTCCTCAGGAAATTCGGTGGGGT 4817
    |||
900 PheAspIleSerLys...G 905
    |||
4818 TTATGTTTGGCAAGAAAGTGTGGAACTGTTAGTCAAGCGGGTGGTTGA 4867
    |||
905 uThrPheGlyArg...LysLeuLysAsnSerArgLeuArgValPheSerA 921
    |||
4868 ATCGTTTTCTCGTTTGAAGCGGAACCCGCTCGTTTAACTCCTTCGCTA 4917
    |||
921 rg... 921
    |||
4918 ATTTGTTGCTGCGAGGCTCTCTTGGGGTTCCTTCGATGACACTGTTTTT 4967
    |||
922 ...AlaIleValGluAspSerIleLysValMetLysAlaMetLysThrG 937
    |||

```



```
8114 GGGTTCTCAGAAATATGATATGCTTTGATTTCTTAATCAACGCTCTTTT. 8162
1880 hrLysGlyLysAspValAspAlaValAlaLeuGlyAspAsnLeuPhe 1896
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
8163 .....GTTGCCGCAATTCGTCTTACGT 8186
1897 ValCysAsp.....AspLeuValPheHisAspAlaLeuAsnLeuIl 1911
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
8187 GCGGTGGATGTTAAGCATTTGACTTTTACCAATGAGGAACATAGCGCTTAT 8236
1911 cGlyAlaLeuLysValAlaAArgCysGlyMetValGlyGluSerPheLys 1928
      | .....
8237 C..... 8237
1928 erPheGluTyrLysCysTyrAsnAlaProProGlyGlyGlyLysThr 1944
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
8238 .....GTTTATGAAGCTCTCCCGGTGGTGAAGACTCAT 8273
1945 ThrLeuValAspGluPhe.....ValLysSerProAsnSerTh 1957
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
8274 TCTTTGGTGAATTCGTACCGCGGATTTATTCGCTTAAAGTA.....TCTTG 8317
1957 rAlaThrIleThrAlaAsnValGlySerSerGluAspIleAsnMetAlaV 1974
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
8318 TTTAGTAGTACCGCTAACAGAAATCTCAACCGGATATCTCAGCGCA 8367
1974 aLysLysArgAspProAsnLeuGlyLeuAsnSerAlaThr..... 1988
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
8368 TATCTAAT.....GAATTGATGGCGAGGAAGCTAGCTGAAGTAT 8408
1989 .....ThrValAsnSerArgValValAs 1996
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
8409 GTTACCGATGCAGCTTCCAGAGTGTCTACTGTAGATTTCTTAAATCAA 8458
1996 nPheIleValArgGlyMetTyrLysArgValLeuValAspGluValHisM 2013
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
8459 TCACCTG...CGACTACGACTCAACTGCTGTTATCAGCAGCTGCTTCA 8505
2013 etMetHisGlnGlyLeuLeuGlnLeuGlyValPheAlaThrGlyAlaSer 2029
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
8506 TGGTTTCATGCTGGGCAATTCGAGCCGTTGAGTTCACGCTTGTAAGA 8555
2030 GluGlyLeuPhePheGlyAsnGlnIleProPheIleAsnArgG1 2046
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
8556 GCGGTGGTTCCTTTGTTGATAGTAAGCAAAATCCATTACATTCACCGAAA 8605
2046 uLysVal.....PheArgMetAspCysAlaValPheValProLysL 2060
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
8606 GATCTCGGGGTTCCCTTTGTTGAGATATTGATCGGCTTATT...CAAC 8652
2060 ysGluSerValValTyrThrSerLysSerTyrArgCysProLeuAspVal 2076
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
8653 CGGAACATCGTATTACGCGCAAGTGCTCTATAGTGTCTTGGACATT 8702
2077 CysTyrLeuLeuSerSerMetThrValArgGlyThrGluLysCysTyrPr 2093
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
8703 TCCGAGTGGCTTCCGACTT.....TACCC 8728
2093 oGluLysValValSer.....GlyLysAspLysProV 2104
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
8729 GCGACACGCTACTGCTAATAGGAGGATTCGGGAAGTCG..... 8771
2104 alValArgSerLeuSerLysArgProIleGlyThrThrAspValAla 2120
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
8772 .....AGTGTGTCATACAGAAATCAATGGTCCGATGACGTACCT 8813
2121 GluIleAsnAlaAspValTyrLeuCysMetThrGlnLeuGluLysSerAs 2137
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
8814 TACGACAAGCGCGCAAGTACATTGTGTACACTCAGGCTGAAAAAGACGA 8863
2137 pMetLysArgSerLeu.....LysGlyLysGlyLysGluThrP 2150
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
8864 TTTGCAAAAACACCTCGGTGCTTAACCGTGGGTTCGGAACAGTAGTAC 8913
```

```
2150 roValMet...ThrValHisGluAlaGlnGlyLysThrPheSerAspVal 2165
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
8914 CAATTGTAATACCGTTCATGAGTTACAAGGAGACACCTATAAAGGTA 8963
2166 ValLeuPheArgThrLysLysAlaAspAspSerLeuPheThrLysGlnPr 2182
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
8964 GGTCTGGTGAGTTAAGTATCAGGAGGATACCCGTTCTCTTCAAGAA 9013
2182 oHisIleLeuValGlyLeuSerArgHisThrArgSerLeuValTyrAlaA 2199
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
9014 TCATATTGTTGTTGCTTGCATAGACATGTTGATTCTTTGTTTATTCG 9063
2199 laLeuSerSerLysLeuAspLysValGlyThrTyrIleSerAspAla 2215
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
9064 TACTTACTAGTAGGAGGTACGACGATACGCACTAACATAGTAGGCG 9113
seq_name: gb_v12:LCCVORFS
seq_documentation_block:
LOCUS LCCVORFS 16934 bp RNA VRL 04-DEC-1998
DEFINITION Little cherry closterovirus ORF's 1-6 RNA.
ACCESSION Y10237
VERSION Y10237.1 GI:2398667
KEYWORDS coat protein; helicase; HSP70 homologue; methyltransferase.
SOURCE Little cherry closterovirus.
ORGANISM
VIRUSES; ssRNA positive-strand viruses, no DNA stage;
Closteroviridae; Closterovirus.
REFERENCE
1 (bases 1 to 16934)
Jelkmann,W.
Direct Submission
Submitted (20-DEC-1996) W. Jelkmann, Federal Biological Research
Centre for, Agricultural & Forestry, Inst. for Plant Protect. in
Fruit Crops, Schwabenheimer Str. 101, D-69221 Dossenheim, FRG
2 (bases 1 to 16934)
Jelkmann,W., Fechtner,B. and Agranovsky,A.A.
Complete genome structure and phylogenetic analysis of little
cherry virus, a mealybug-transmissible closterovirus
J. Gen. Virol. 78 (Pt 8), 2067-2071 (1997)
97410293
FEATURES
Location/Qualifiers
1..16934
/organism="Little cherry closterovirus"
/strain="UW2"
/db_xref="taxon:53180"
77..6982
/note="ORF1a"
/codon_start=1
/product="methyltransferase (MT) and helicase (HEL)
domains"
/protein_id="CAA71285.1"
/db_xref="GI:2398668"
/db_xref="SPTREMBL:O40959"
/translation="MAFLSNVVEVTLFTCISPNIKISKTKVPTPEYLKLMKREIERS
LALANARSPRISVPVAGPTPTAMTTFSESSDISRLGKLPKFLKNSREAVLDKQ
MLTRNPENATGFKCLPKLAKPTSTIRSKTPAVRTAKLSFNDFSDAGSINKPAVS
KLELSQAALAESPEARVGAAPVQSSLHWTLNADIIRSNLPRGLVCEIYVSLENKL
LKLVSFDKNFPELLIAEVNIPNNNGKVPHTVNFVIAPSAGVDSKLPVRSYLLNP
HGHELSLLSFLIHTNPFEDIVDALNLFKPDRLHALHYRDFCINHLIDVGNILNF
SGTLRRPASTGYVRKPLVRAHQAKAFISFSEFSALEKQKASATAIKFGDF
EVFDLREPNNVTKSYAALTSNLPVDDRILKLNIPVSEVKTEPCSTMNPSNPKSEA
KRAVELSSRATSDSGEILYSKTFNOLLNPTGCT11RHFGINGVCRSFTLNRSQVRFN
ILNRNAKIMFNLSLKHKKFFYVDGTCRTPSGKLSEYDNGFCWLOAFWAFGIPTIF
VEFIPNLVNSCLLQAGPKSKLHMHTKNTALYHFDPNKINNIIYHANLNFVLVSAMNE
NKDILESNDILIEVGTFOVIAKAYPKSDNQATNLNARVGRINTCDCRNDKMLIF
VHLSTRQKIKVCDLFLPHLKFTQSSHPVENVVTVTSNFSVLYKMQGRHPVDFGG
NIGTVINSECDVHICNPVADSRDAKRHDVNGFLAKSVGSNNISVKNKLAOCHNHK
SORAVMEVYDMLTLEMCOAMLAHGTIRLDFILLPCDLLEDFNTIIFDGGCKITKD
DOKVYIYGDAEAYTHDLNLENMTDNLNLCVCGTAPFKTKLETSGPPRHFSLTKLE
TFPSKIEFLTMYDKCKENKMLVKVPMRNINGMTVLHDLIVDRSMLNMLIEYCANCE
NPNRKGFEHIMSQRKNYVIYNGKVIHDDQVDIPLOLPGFVGCLSEGIRLAERVH
YLAKYSYQYVYAGIFNCMIYSRQLFLRFKTVCEYKVLDLMDLFLGCEWITTTARNSG
```

```

TKIPDISAVVVKQSLISADGDVHSIFTEPFEXYQORCADLEQHASFEVDTASELSDA
IDSGMGVSDLKIPKFTIKLAELGFGERNVLRFSAKLYSPDWLRFGSNDNSVLRRV
IGYKSLSTLKNILLSPYLLIKWICDCGKASPKTFVSGLLNVFQTSORINFLVK
IALDSDEDFQNEVAAFRHDLAEADPTTGKVTNFQVVDSDTKKKIISIIIDEEL
SYERIGDIVDQVDRKEILEIASADPTITKVNNGHLDDALEDNVRHLKIDIQLRKE
VLRVDEKIDALESGAGKSKLCKDIENYLDRNLNFKHYITSKGDFDYDLHSEF
EEVKRGSAILCALVHDDLSDIYDNGVLVLPKNSAREVLYARLAKMKELYRNPMTT
FIPSMRFLSGIKNKMISHICDAERQAFIDFKCLDLYLVISCTATAQLFGGNECLIRLI
LSPILSALTKKSWTYIFGEHOGIISTIIASLLGGGNSVNMALIVSGSYFHFDPYFC
NKMINSFKFYDVIYIARHNEFTTAKTIVGLITSPGVAVLTGVALSGINPFAAMFF
FAASHYKNFVDGIVRVSIAIATGDTLLHLKPVGALKAKINVKRMFKNSNITSNS
TRVNEVNDGHDIANDTNKNSETSNIAYEEVEEQSDDEIDVIDYMKRVKDKMPKIEY
IDQSEPCVKQALTNOLLKSDYNPNASSYSTGAVDHLVPDGYQSVDFDKNKHGP
DISVOLCELPKCELMRNPASKNIVFOQTVNEVENSQEFVQIESLILLNFKKLET
TLDKFSGKYYRQLQTLVNDSTQPMYKQGQDVPLSLIHSDDVPOTVIEYSTSRKLA
RFDNSGDEVQTTTEFAAGFSNKKLISMDFKFSQSANIIDSDRLSMKTVNKPFGSGKT
TQISOMMFDCHIRKTCIALTSTKYGNKELIEKLNLIIVGAKPKVKTVDAFNMHGKK
AIVDELYVDECYMHSAGAILMALSNVDFTNCKFGDENQIPIYSVRLTGFALFATVVF
KYCVTHDSISYRCPADVCLLSHATDRAGNMXTNGVYAVGNQIRTMHNLNALTGVA
NIPUREDPVYAAFTQGEKDELREAKKNIKTIVNELQGTGFKNLIVFRLRLHNWLYN
DTRQFIRAISRHTETFNFCFSNAMNDSISNAINKLANTIENNVVAQGFQCV"
77..8495
/notes="putative fusion product expressed by ribosomal
frameshifting in 0/+1 configuration"
6951..8498
/notes="ORF1b"
/codon_start=1
/product="RNA polymerase (POL)"
/protein_id="CAA71293.1"
/db_xref="GI:2398669"
/db_xref="SPTREMBL:O40960"
/translation="MPNLGSDSVEDEYSLVWEDRFPNVFPARPSLSSVHAVNDFMDLT
FNGLSSVDYLRTRUFEXYSDIELPCELDLLITPTKSPYQAVNCKVPILIGKGRSRP
DTWKVYLSKRNFAAPRLNEDISIDETANRLTSLMRCMEPSLAQYFDVPEPDVN
KINSWLTDRKRYGNVLRFGDNGDWTTLFNLKMWGDMKPKLDMGSGVSEYSPASN
LIYQMTLMNFPFSPITFIEGRKYICVIGDKVILFSGMLDEMGLDLSRLCYPLNSYN
FVEVDFSKDSQGHVIKLYEELVYKIEKSPNLDNEKLSYEYCRASATCGVNLIDLY
CQRRTGSPNHLNSLATLAMLASYDLDDIDLIIVGDDSLIISRKIEENKLEININ
DQGMADKFLANPVPYFCKFIQVDNRIRLVDPPVRFPEKLTPTVLOLEHLEITRE
RFTSYRDMGYSYFDENVIIAVERFVSLKYNTPMGSGYAACYIHCLLSSFRNFLTIFG
DTSITQL"
8503..8613
/notes="ORF2"
/codon_start=1
/product="unknown"
/protein_id="CAA71286.1"
/db_xref="GI:2398670"
/db_xref="SPTREMBL:O40961"
/translation="MVVLVWVCLISVWCVFFVSPNSIGLEKFSRFGFSNVC"
8603..10459
/notes="ORF3"
/codon_start=1
/product="HSP70 homologue"
/protein_id="CAA71287.1"
/db_xref="GI:2398671"
/db_xref="SPTREMBL:Q98159"
/translation="MSVKGVIDFGTFTSITGFVNGSEVSVLLVDKTPFVPSVLAIFDG
NNVLVGEQAKVISRFQVYDLKRWVGSEKNFNSLKEKISPKYDIFKNNDYI
SGVGTQRLIYRILSYLYWLKILNFDOLKLVNEVINSVPADYTRQRYMRSIV
NLIGIPVRIINEPSAAMHOLFNPKENNVFVDFGCGTFQVSVIKKHGKIVSITD
AGDPLGGRDIDKALAEFISQKTGHSCDDDLFINHLKEVTFGCVSKNKFVONSGLIRS
VDNLNINSVAQFIPNSKLLDSDVISRNHISKALVCWGMGDSMTDGTYSNVTKYCD
SKGLKVFNSHRSVAYGCAILHDFDSKDFIYVDVNSHGLIDVGHFVPRVIMRKP
LSIYTSFDTNLKLYSTAACVYEGDSLNFLENDLLVNATFFTFDESDGLSGVRLTY
SYDIEGLSVSISSTDGTFKFNLENKVDNSELVKLDLIQTLQSSSAYISVIVLSKY
WGINPELKLAIERPYLKDFDKNGGLEEYVDLRNEIGAFAVDSDRLVRLVLLFIL
DGVKPYLYSNCKHFSGDLVLLTLLILHLKIYSENFILIRKNKGRDKQKFF"
10431..11984
/notes="ORF4"
/codon_start=1
/product="unknown"
/protein_id="CAA71288.1"
/db_xref="GI:2398672"

```

misc_feature

CDS

CDS

```

/db_xref="SPTREMBL:Q98160"
/translation="MEIESKNSRILSALIFKHDISEIRGFKNLSKDKINKKVPFHG
VSGNNEEIKFNGTNDKNLVIKDDDIISIIFLCIIYTFEFDRRFVDSKAYPIENTLC
FYDHLDTIELENFKGESLLSYCDKHPNVCCKFTIDDIKEKFSNLDSDLYFSYRLSN
CUGRLTSVLDAKSTDFSVLRKPTLLSKSYVDQSDTLRCKMHDHYEISSESLL
PLYQSOLWCSLIYLLLPVDERDKLYNNEIFLGLCLYSLIGDYINKEKDSLKIRTSILD
FIDDLHLTLKLFSEGFEDFKVKYIKSEESFASRQLFEINELQMLLESKAKOVNASY
AQSLSQDLVNMENIVSYFRIYFNISHPELAVALTVPFFSPMSTSPNIRKRTPEKFF
IYSDDGREKIGKIRSNHFVYIDSKNELTFRDRDNRNIRLFCSKRANLAI RVNEKF
RFQPTVMQCPKVPYMKIDFYKGLKATILLTKEENSLAHLKLYTDYVSDNSKFSKDK
YFDYLMKPT"
12064..13278
/notes="ORF5"
/codon_start=1
/product="putative viral coat protein"
/protein_id="CAA71289.1"
/db_xref="GI:2398673"
/db_xref="SPTREMBL:O40962"
/translation="MAEAFKASFDRLVOLLPIVPIWLI PNPSRSLOQKSDVGEDTA
TKISIINAHIIYKEDLNKIAOLDVGTSAPEVKLLAEVRKCRGKVOISYTNVVKNF
RDYQRLISKPVADRNTDEINLNFQDLTLLENVKSTLVELTELKFLSSFNITNIPEN
ERNQFKNLKEFVDSDEIEEENVRHSDEPVSRLNFKDLDFQSVKNIKCOASNELTD
ELTDLFLKLEEHFRDIPFKSSGVLTPAETWSALVSYFTSLYEQSTTKLADSKDA
LNSFSVGKTIKYEQPGVPTKLTTEHRSKHGVENAEERREKKEYNAIQNALAAGTVS
SERLAAGKAAPNKRKGVSDATPMKHAHMTPEQSVQLASTDTATDTDTSSSKGALHV
SOVLGPRRGRRN"
13284..15278
/notes="ORF6"
/codon_start=1
/product="putative viral coat protein duplicate (CPD)"
/protein_id="CAA71290.1"
/db_xref="GI:2398674"
/db_xref="SPTREMBL:Q98162"
/translation="MNFYDIYITSRIPNLSVNIKLSDTCTCAVDKTYVYVIGDSN
DNVKTLLSPDHI SLNMRPRIELVNGNLLNQSEYTLNKTDDHDSNLDPRNFTYS
ANGFNMTLSPVHISLVNRYELIYNGSLKTDIAGVAFENKNNLLRILDERVFSALIKS
PSHKTEISLDRAPYLGSLYNDGPAHKSMAVYKNKIENIYVODLIPSNFIRKNLND
GUSNINLEVNNNNNCLRVYDVIKCSDESHGVTLEQLTQNTSEALNVIFQWYIGVG
GNILBEVLRKNSCFGECKNNGGYSKLTDIRYHGKTDYISGIIIFGVYIDRS
QNVFTINGVIVDMKNELLNAYVEFGFQFGATEKYNEFKRKRSDNAIYRIL
RYPYNLERPYTSDMTNHHNNVLTNSQVFLSEVNGVRDVENTPOSRPTLDSSRNKE
TKRPRVESDSDDEPVLTAISGETKLGDPPIELKQITNOIADNSPESKPYVWIFE
YAKDCDQVITTIKGLQTSITDAKLVVFWMAICCGSTIESVHDKHTSLVNLNNKKI
YLRATCFEKKKPNKINLFRYLRSTVEVLELRIGTLPSPSFGRAIMLIGIPKQFAFL
ACDFWNFDEMLTTQHEHEVIGNLKSVSQTSQRIQFVETIR"
15426..15971
/notes="ORF7"
/codon_start=1
/product="unknown"
/protein_id="CAA71291.1"
/db_xref="GI:2398675"
/db_xref="SPTREMBL:Q98163"
/translation="MLPQVLECKRFEARRDETLETIGSRKMFSEILEIDETVNE
FKELFLMRIPKLSVQYCDDFILVFNILKNFFDKTLINKEKDIGEMDFKGTWVKLKM
SEELTEILRNKQPSNNLKFSUKVSECLKGFTTLLSSMFWYNRGRNRAYAKAITLNL
QLSIEHSGISNGELAYNLVVG"
16008..16727
/notes="ORF8"
/codon_start=1
/product="unknown"
/protein_id="CAA71292.1"
/db_xref="GI:2398676"
/db_xref="SPTREMBL:O40963"
/translation="MFKLENKFLSKRKLKMANSTFFSNSMDTVVNDILAFPLVLY
NSINRTPDKFCIYDLENKFSYLNLTLEDKNIFYSNSSTVSWDDGNNSVNEH
DADGHLGIPLYEAFICLFDILNFRULKHLQTRNLIAFENVQLLAKGKNMLNDNI
NKLFEKDDGIEFNFYHSAFVKLVANPLISLKKLKESTGWEDGMDLLSDVMEILHQNV
IETFASEKDDMHIMSTILHIS"

```

BASE COUNT 5433 a 2681 c 3292 g 5528 t
ORIGIN

alignment_scores:

Quality:	721.00	Length:	2432
Ratio:	0.665	Gaps:	91
Percent Similarity:	44.572	Percent Identity:	19.285

alignment block:

US-09-301-906-15 x LCCVORES

Align seq 1/1 to: LCCVORFS from: 1 to: 16934

58 ArgThrPheValProProValValGlyPheAlaArgGlnTyrAlaIva 84
||| |||
1085 CGTAGCTTCGTCCGTACGTACTAGGTATGTCGT 1123
84 IvalSerGlySerValSerAlaLeuArgGlyAspGlyLysValLeuM 101
..... AAACCAAGTTTATC 1136
1124

101 et.....GluAlaArgThrSerThrSerAlaThrSerAspValSerAsp 115
 :::|::| :::|::| :::|::| :::|::|
1137 GCGCACATCAGGCTGCTCAGACTGC AAAAATTTTCAGTTCATCTCTGAA 1186

```

116 pAcAspvAlvalPheGlu.....AlavalSerAsnAlaLeuLeuVa 129
    |||:::|||||      |||:::|||||      ::
1187 TTTTCTATTGCCTTGAGAAACCAAAAAGCAGCATCCACGGCATTCAGCCAT 1236

```

129 lValHisTyrHisArgValValProTyrAlaProValLysArgGluGlnP 146
:
1237 T.....AAATTCGGTGATTTCGAAGTTTTGCACCTGAGAGAAC 1274

146 roLysProAlaValLysGlnAspGluGlnLysProLysArgGlnAlaSer 162
 || ::::||||: ||| :::||
 1275 CTGATAATGTTATCAAGTCATACGGCTCAG.....GCTCTGACTTCA 1315

163 H1StrpAlaValLysProThrAlaValGlyValHisValProLeuProLy 179
::: ||| :::::||||: ::
1316 AACTTGCCTGTTGACGATCGGATATTGAAATTTGAACATCCTGTTTCGTC 1365

179 sLysGlnGluAlaLeuGluProAlaGlnSerValProGlnGlnSerLeuG 196
::: ::::: |||:: |::: :
1366 CGAAGTTTAA GACTTGAGTGCCTCCATCGACCATGAATCCGTCTAATCCGAAAT 1415

196 luGluLysAlaAlaLeuThrPheGlyLeuPhePheSerLysGlyGlyGly 212
 :
 :
 1416 CTGAAATTGCTAAGCGACCGGTTGAGCTG...TCTTCTCGCGGACTAGT 1462

213 AspGluSerAspAlaValIleLeuArgLysGlyLysLeuPheAsnArgAl 229
||||:|||| :: :||| |||||:::
1463 GATTCGCTCGTGAGATTAT.....TCAAAACCTTTAATCAACT 1503

229 aLeuAsnValPro..... 233
 |||||::|||
 1504 ACTGAATATCCAACTGGTGGGACTATAATAAGACATTTTCATCAATGGTG 1553

234IleAspVal 236
1554 TCTGCAGGTCATTTACTCTTAACCGCAGCGATGGGCGTGTTTTTAACATT 1603

237 LysAsnThrPheValTrpAlaLysIleTrpAspGluAlaSerArgArg 253
 |||::: ||||||| ::: ||| ::::
 1604 CTCACAATAGAAATGCTGCTAAGATTATGTTCAATCTCTCGCTGAACA 1653

253 gGlyTyrPheTyrValLysAspArgAlaValLysPhePheProIleValA 270
 : |||||
 1654 TAAGAAAATTTTATGTT..... 1669

270 rgGlyArgAlaThrIleGluAspPheIleValAsnThrAlaProGlyCys 286
 1670GACGGCTATTGCAGAACACCATCTGGA... 1696

287 AspValAlaLeuProArgIleGluLeuTrpSerMetArgGluArgAlaPh 303
1697AAATTATT 1704

2470 GTCATCGCGTGTATGGTTGAAGTTTACGATATGACTCTCACTGAAA 2519
 602 alalacGluAlaMetAspLysGlyGlyAlaLeuValPheAspIleAlaLeu 618
 2520 TGTGTCAGGCCATGTGGCACATGGAACATCATCAGATTAGACTTTATATTG 2569
 619 MetPheProValGluLeuLeuTyrGlyAsnGlyGluValTyrLeuGluG1 635
 2570 TTTATACCTGGTGACCTTCTTCAGGATTCTCAATACAATAACAATTTTCGA 2619
 635 uLeuAspThrLeuValLysArgGlyGlyAspTyrLeuAlaTyrAsnValG 652
 2620 CGGTGGTTGTAAAGTACAGAAAGATGATGACAAGTTTATTACTATTATTCG 2669
 652 IyGlnCysGlyGluMetTyrGluHisSerPheSerAsnValSerGlyPhe 668
 2670 GTGATGCTGCAGAAGCATATACCCACGACCTGGAACAATCTCCGTAAATATT 2719
 669 PheThrPheSerTyrValArgThrSerSerGlyAsnValPheLysLeuG1 685
 2720 ATGACTGATAATCTTGT...TGTGTTGATGGTACAGCCTTCAGAAAC 2766
 685 uTyrGluGlyTyrArgCysGlyTyrHisHisLeuThrMetCysArgAlaG 702
 2767 TTTGGAACACTAGTTACGGTCGATTTGGTCATTTCTCAATTAACCAAGTTGG 2816
 702 InLysSerProGlyThrGluValThrTyrArgSerLeuValProSerPhe 718
 2817 AAACCTTTCCTAGTGGTAAATGNAATTTCTACCATGTATGATAGTGT 2866
 719 ...ValGlyLysSerLeuValPheIleProVal...ValAlaGlyLysSer 733
 2867 GAGAAAGATAAGATCTTAGTCAAAGTGGCGATGAGAAATATCATCGTAT 2916
 733 rValSerPheLysThrIleValLeuAspSerAspPheValAspArgIleT 750
 2917 GGTGACTTTCGATGATTTGATCGTAGACCGCTCAATGTTGATGAATTTAA 2966
 750 ySerTyrAlaLeuAsnThrIleGlyThrPheGluAsnArgThrPheGlu 766
 2967 TAGAATATTCGCAANTTGTGTTGNAATTTTACAGAAAGCGTTGAG 3016
 767 TyrAlaValGlyAlaValArgSerGlnLysThrHisValIleThrGlyLys 783
 3017 CATATAATGCTCTCACTATCGTCTAGGAAGAACTATCTCATTTACAATGG 3066
 783 rArgValValHisSerLysValAspIleSerProAspAspMetTrpGlyL 800
 3067 CAAAGTGTACCAAGTCAAGTTGATATCTCTTTAGATTTTACTCCCTGGAT 3116
 800 euValValAlaValAlaValMetAlaGlnAlaIleLys...AspArgAlaLys 814
 3117 TTTATGGAGTTTGCTATACAGAGGAATCCGCTGGCTGAGAAAGTTCCAC 3166
 815 SerIleArgSerTyrAsnPheIleLysAlaSerGluGlySerLeuAlaG1 831
 3167 TACTTAGCGAAGTATTCTCTACTATCAA... 3193
 831 yValPheLysLeuPheGlnThrValGlyAspCysPheSerAsnAlav 848
 3193 alSerValTyrAlaLysAlaMetValHisAspAsnPheAsnValLeuGlu 864
 3194 ...TACTATGCTCCCGGAANT...TTTAATTGTATG... 3223
 865 ThrLeuMetSerMetProArgAlaPheIleArgLysValProGlySerVa 881
 3224 ...ATTATACCATTCGTCATTTATCTTAGCC... 3253
 881 lValValThrLeuCysThrSerGlyAlaSerAspArgLeuGluLeuArg 898

3254 .TTCAAGACGTGTTTGCTATGAGAAGTCCCTAGATTAAATGGATTTT... 3299

898 lyAlaPheAspIleSerLysGluThrPheGlyArgLysLeuLysAsnSer 914
::: ||| ::: :|||::|
3299TTATTTCGAAGTGAATGCACACTACAACAGCAGCTAAGTCT 3337

915 ArgLeuArgValPhe.....SerArgAlaIleValGluAspSerIl 928
:||||| :: :|||::|
3338 GGGACAAAAATATTCGATATCTCAGCTTCCGTAGTAGTGGTTAAGCAAAGTTT 3387

928 eLysvalMetLysAlaMetLysThrGluAspCysLysProLeuProIleT 945
::|::| :|||::|
3388 ATTGATT.....AGTGCAGCGGG..... 3406

945 hrGluAspSerValTyrrAlaPheIleMetGlyAsnValSerAsnValHis 961
:||||| :|||::|
3407GATGTC CAT 3415

962 CysThrArgAlaGlyLeuLeuGlyCysSerLysAlaThrValValSerSe 978
:||||| :|||::|
3415 3415

978 rValSerLysGlyLeuValAlaArgGlyAlaAlaThrLysAlaPheSerG 995
:||||| :|||::|
3415 3415

995 lyIleThrSerPhePheSerThrGlySerLeuPheTyrrAspArgGlyLeu 1011
:||||| TCTATCTTT..... 3424

3416 3424

1012 ThrGluAspGluArgLeuAspAlaLeuValArgThrGluAsnAlaIleAs 1028
:||||| ACAGAGAAC..... 3433

3425 3433

1028 nSerProValGlyIleLeuGluThrSerArgValAlaValSerLysValV 1045
:||||| :|||::|
3433 3433

1045 alAlaGlyThrLysGluPheTrpSerGluValSerLeuAsnAspPheThr 1061
:||||| TTGAA 3439

3434 3439

1062 ThrPheValLeuArgAsnLysValLeuIleGlyIlePheValAlaSerLe 1078
:| | |
3440 AAATATCAACAAGG..... 3454

1078 uGlyAlaAlaProIleAlaTrpLysTyrrArgArgGlyIleAlaAlaAsnA 1095
:||||| :|||::|
3455TGTGCAGATC 3464

1095 laArgArgTyrrAlaGlySerSerTyrrGluThrLeuSerSerLeuSerSer 1111
:||||| :|||::|
3465 TAGAGCAGCATGCT...TCCGAATATGATACC CGCTCAGAATATCCGAC 3511

1112 GluAlaalaGlyGlyLeuArgGlyLeuThrSerThrValSerGlyGl 1128
:||||| :|||::|
3512 GCAATCACAGTGCAGCATGGTGCTCTCTGACTTAATAAATT..... 3553

1128 ySerLeuValValArgArgGlyPheSerSerAlaValThrValThrArgA 1145
:||||| :|||::|
3554TTTGAGAAATTCATAC..... 3571

1145 laThrValAlaLysArgGlnValProLeuAlaLeuLeuSerPheSerThr 1161
:||||| :|||::|
3572AAATTAGCTGAATTAGTTTTGGAGAG 3598

1162 serTyrrAlaIleSerGlyCysSerMetLeuGlyIleTrpAlaHisAlaLe 1178
:||||| :|||::|
3599 CGAAATGCTCTAAGATTTTCAGCTAAGCTGTACTCTTGTTTCGATGGTT 3648

1178 uproArgHisLeuMetPhePheGlyLeuGlyThrLeuLeuGlyAlaA 1195
:||||| :|||::|
3649 GCACAGTCFAA..... 3658

```

1195 rGalaSerAlaAsnThrTrpLysPheGlyGlyPheSerAsnAsnTrpCys 1211
1369 .....TCCAAATACGATTCT 3673
1212 AlaValProGluValValrPrArgGlyLysSerValSerLeu..... 1226
3674 GTCTTCGGAGAGTCATTGGCTACTTCAAGTCTTTATTTCTACTCTCTAC 3723
1227 .LeuLeuProIleThrLeuGlyValSerLeuIleIleArg..... 1239
3724 TCTGAAGAATATTATTAATTTCTCCTTATCTTAATTAAGTGGATGCTGG 3773
1240 .....GlyLeuLeuAsn.... 1243
3774 ATTGTCGGAAGCTACCTTTAAGACGTTCTGTTTCAGGATTATTGAATGTT 3823
1244 .....AspThrIleProGlnLeuAlaValrValProProVa 1255
3824 TTTCAACATTTTCGGATAGAATAAATTTTGTAGTAAAGATTGCATTAGA 3873
1255 lGluGlyArgAsnValTyAspGluThrLeuArgTyTyArgAspPheA 1272
3874 CTCTGATGAAGATTCCAAATAGAGTTGTTGCAGCCTTTAGAAATCATG 3923
1272 sp.....TyrAspGluGlyAlaGlyProSerGlyThrGlnHisGlu 1285
3924 ATTTAGAACCTTTTCACGATACCGTTGGCAAGTGTTTACAAAAATTTT 3973
1286 AlaValProGlyAspAspAsnAspGlySerThrSerValSer..... 1300
3974 GATGTCGTTGTCTCTGATGACTTCATGAAGAAGAAGATCATATCAATCAT 4023
1301 .....SerTyArgp.....ValValThrAsnValArgAspV 1311
4024 TGAAGATGAACTCAGCTATGAAGAGGCAATGATGTCGACGTCAAGATG 4073
1311 al..... 1311
4074 TCAAAAAGGAGATTCTAGAAATAGCCAGTTCTGCAGACACTCCAACAATA 4123
1312 ...GlyIleSerThrAsnGlyGluValThrGlyGluGluGluThrHisSe 1327
4124 ACTAAGTGAATAAATCTGGGCCACTTGGATGATGCATTAGAGAAATGACGT 4173
1327 rProArgSerValGlnTyThrTyValGluGluValAlaProSerA 1344
4174 CAGGCATTTAGTCTTGACATTCAACTCAGGAAGAAGTGTG.....A 4217
1344 laAlaValAlaGluArgGlnGlyAspProSerGlySerGlyThrAlaAsp 1360
4218 GCATGTTTCACCAAGACCTTAATGAAGCTTTAGAAATCAGGTGCGGGC 4267
1361 AlaMetAlaPheValGluSerVal..... 1368
4268 AAATCGAAGTTATGTGATAAGATTTTAACTATATCTTAGATCGCTTAA 4317
1369 .....LysLysGlyValasp.....AspV 1375
4318 CTTATTTAAGCAATTACATCACATCTAAGTTTGGATGGGATTTTATGATA 4367
1375 aLPheHisGlnGlnSerSerGlyGluThrAlaArgGluValGluValAsp 1391
4368 TTTTTCAT.....GAATCTTCGAGGAAGTTAAAGTGGCG 4402
1392 GlyLysGlyLeuLeu.....ProGluSerValValGlyGluAlaProTh 1406
4403 AAGTCCGCCATCTCTGCTTATCATGATGATTTGTTCAGACATTTATAC 4452
1406 rGlnGluArgGlyArgAlaAlaAspGlyAsnThrAlaGlnThrAlaValA 1423
4453 T.....GACATGGTTTGGTTTAAACCCTTGA 4481
1423 snGluGlyAspArgGluProValGlnSerSerLeuValSerSerProGln 1439
4482 ATAAGTCAGCTCCGAAGTCTTTATGCCCCGACTGATTGCTAAG..... 4525
1440 AlaAspIleProLysValThrGlnSerGluValHisAlaGlnLysGluVa 1456
4526 .....ATGAAGTTTAT 4536
1456 lLysGlnGluValProLeuAlaThrValSerGlyAlaThrProIleValA 1473
4537 TTACAGGAACCTTTATGACAACTTTTATC..... 4564
1473 spLulysProAlaProSerValThr...ThrArgGlyValLysIleIle 1488
4565 .....CCACATGCGGATTCTGAGTGGGATA..... 4591
1489 AspLysGlyLysAlaValAlaHisVal.....AlaGluLysLysGlnVa 1503
4592 ...AAAAATAAAATATGTCTCACATTTGTGACGCTGAAAGAGGCAAGC 4638
1503 lGlnValGluGlnProLysGlnArgSerLeuThrIleAsnGluGlyLysA 1520
4639 TATCTTTGACAAATGTTTAGATGCCACTCTAACTGTG..... 4675
1520 laGlyLysGlnLeuCysMetPheArgThrCysSerCysGlyValGlnLeu 1536
4676 .....TCATGCACCTGCTACAGCTCAACTT 4699
1537 AspValTyAsnGluAlaThrIleAlaThrArgPheSerAsnAlaPheTh 1553
4700 .....TTTGAGGGAATTTCTG 4716
1553 rPheValAspAsnLeuLysGlyArgSerAlaValPhePheSerLysLeuG 1570
4717 TCTTATTAGATTGATA.....CTTTCACCTATTTTAAAGTGTCTACCTTA 4760
1570 lyGluGlyTyThrTyAsnGlyGlySerHisValSerSerGlyTyTrpPro 1586
4761 AGAAATCTTGGACATATTTTGGCGAACAT..... 4792
1587 ArgAlaLeuGluAspIleLeuThrAlaIleLysTyProSerValPhe... 1602
4793 .....GATGGTATTATCAGTACTATTATTGCTTCGACTTTATTAGG 4833
1603 ....AspHisCysLeuValGlnLysTyLysMetGlyGlyGlyValProp 1618
4834 TGGTGGGAATATAGTGTAAATAAATGCAATAGTTCTGTTTCATACT 4883
1618 heHisAlaAspAspGluGluCys.....TyrProSerAsp 1629
4884 TCCATTTCAAGCCTTACTTTTGCATAAGCAATGAAGTATAATCTTTC 4933
1630 AsnProIleLeuThrValAsnLeuValGlyLysAlaAsnPheSerThrLy 1646
4934 AAGCAATACGTTTACA...GATTATATTGCTGTCGTCACAACCTTTTACAAC 4980
1646 scysarglys..... 1649
4981 GGCAAAAACAACAGCTGGTGGTTGATTACGTCACCAGTTGGTGTGTGTC 5030
1649 ..... 1649
5031 TTACTGTTGGAGCTTTATCTGGTATCAATCCGTTTGCAGCTATGTTCTTC 5080
1650 .....GlyGlyLysValMetVa 1655
5081 TTTGCAGCAGCTAGCCACTATATAAATTTTATGATGGTATTGTTCCAGT 5130
1655 lIleAsn.....ValAlaSerGlyAsp.....TyrPheLeuMetProC 1668
5131 TTCAAAACATCGCTATTGCAACCGGGATACCTTACTGACACCAAGCCCTG 5180
1668 ysGlyPheGlnArgThrHisLeuHisSerValAsnSerIleAspGluGly 1684

```



```
NSVSYGDAEYLAASGEPGSFYKDLRWVGCTAKNYOTYLHLKSPPKYKTVKEKFTKS
VPVYLSPLNDLGLNVALPLLIAASYAKSILSDAERVFNVSCTGVICSVPAGYNTLQR
AFTQOSISMGSYSCVYIINERSAAYSTPLKLSADKYLAVIDFGGGTFDVSIVSVRL
PTFAVRSSGDMNLGGDRIDDKKLSDIYEMADFIPQKELNVSSLEKALSLOTDPVKYT
VTHMGSSITIDQTAUREIASVFINRTIDILTQVKVKSMPESQSLKLVVVGSSYL
PGLDITLATVPFVSGIVPEGEOKLTLRKCVSTSYOARMEGDEYKVFERNERYAASVSL
SADSVVAAGSPPIPFEGEOKLTLRKCVSTSYOARMEGDEYKVFERNERYAASVSL
FTLGVNHVPNDVEMTLVTKVDSMKVEFYLLKPGSELVNVQGTSRVDYTGMPHPTRN
LYKLSDVNSAALVALTLTRKREKFLTLFTDLLADLRRITASLSEYSKKYPITR
NDIDVSSRMGIIVSVKVLRGSDLERIPL"
13712..15319
/note="ORF6"
/codon_start=1
/product="p61 protein"
/protein_id="CAA77166.1"
/db_xref="GI:4239720"
/translation="MSSHVMGSLFRKFPYGEAIWKKEYLSESTRSDFERNVSLDHTLSS
GVVVRSLNAPQCTFENELALYNSVVIINDFVELTCMPLKSLMTGIEDRKVPDELI
SVDPHEVGCRTLLNDVESYLIISRGESADVTAVESHWSCLNSCGRLIASSTREIDAKYI
VETKNFSNVSGVTKLETYLSYICISLYKHKMDDAYFNLTIPMFCNLVVLASLGL
FYKHTDNPLLTGLMILEFCLENKYIYSTFKVNLONRLFKSVLPVLVWIDISEPDD
PVDERLIPDPTDFVLDLPKLNIDTMVVVGNQIROLEYVESDALDLSQHVDIRL
AADNPLVLRUMAGMFIYGVYRCADRVERPTLFLPKQLLRQDDGESCRLMSA
VEALFNLVQKVNROIIVRRQPMGRKSHVALRLYRNIGLRFPPISVRLPVHGVLYVD
FYKRPDVAVTADELSRLRLRSVDVMCKDRVSTITPPFNRLRGSSRTFRGCARG
ASSRMSRDVATSGFNLPHGRLYDS"
15294..16016
/note="ORF7"
/codon_start=1
/product="p27 protein"
/protein_id="CAA77167.1"
/db_xref="GI:4239721"
/translation="MAGYTLTLPNTDDKEMDPVSAAPGKYPDVIEKFPVANKRSVDALIE
GVISKLPTNSYEDSTKFTGEHLKYVMVMDAPLLENYKTKTDEDLHLVHAMIQRKLC
TTSSTKTRKDKCISVYQGSRYKLMKDVFPFIISKFTHRETPNALRYACTFEE
LHLCMARLPDLYENKRTTKAGTPhLKGYSADFLSGSLPCYSEHERGIIILRASESML
AROGYBEATELLNRLDKYL"
16116..16787
/note="ORF7"
/codon_start=1
/product="p25 protein"
/protein_id="CAA77168.1"
/db_xref="GI:4239722"
/translation="MDDETLLKNKEMKEGDDVVAESSEFGSVNLHIDPILTMND
VRQSTQONALNRLDLFALKGYPNLPDKDKDFHIAAMLYRLAVKSSSGSDDDTGT
ITYTREGVEVDSLKMTDVIYNSKGIGINTALNRVWGRNTDALYLAFCRQNRNLSYG
GRPLDAGTPAGYHYLCADFAGLTDECAVYIQAKEQLLKKRGADEVVVTNVRQLG
KFNTR"
16753..17256
/note="ORF8"
/codon_start=1
/product="p18 protein"
/protein_id="CAA77169.1"
/db_xref="GI:4239723"
/translation="MSSGLNLTHTVLDLLRSDFLSGWWSFVNVGDIIVRFALHPVN
EDMUNSFSAISNCTIIAGSALLKDNITVDRLESMLPLAYLLKLAKTTTITCFTFMSNK
VLFTTKEPSLCAITSDRVLFVKVINGANVDDSRCCGCSIMFYNNGTFTQNGLTRCNLL
17290..17649
/note="ORF9"
/codon_start=1
/product="p13 protein"
/protein_id="CAA77170.1"
/db_xref="GI:4239724"
/translation="MSIRRVMLKVIAVITVLMCKEPSISEGYNALMNDDFKFDTHF
SNVSYAKCYDLANFDLDFLRIVILPSGGTANESRSDRTNVSEIVESHVSDRDMHI
```

alignment_scores:
Quality: 711.00 Length: 2407
Ratio: 0.676 Gaps: 102
Percent Similarity: 43.664 Percent Identity: 20.482

alignment_block:

```
US-09-301-906-15 x CITV18420
Align seg 1/1 to: CITV18420 from: 1 to: 19259
298 MetArgGluArgAlaPheValCysThrThrLysGlyTrpCysTrpPheAs 314
:::||||: ||| ||||| ::| ||| ::| ||| ::| ||| ::| |||
2779 ATTTCGTACATTCTGAGGTGTGTTCAGTATGGGAAGGTTTCTTTTAG 2828
314 nAsnGlu..... 317
2829 GAGGAGCGTTAACTTAGGGCGCTTCCCTTATGTTTTCGAGGTACAGCACC 2878
317 rgLeuArgGlyGluIleTyrArgArgCysPheSerSer..... 331
|||||: ::|:||||: ::|:||||: ::|:||||: ::|:||||:
2879 GTCTTGAG...CGGTGTATGTTAAGCGGCTGTGTAGTACAGCGTTAGA 2925
332 .....PheSerIleGlyPheLeuMetHisLeuGlyPheArg...SerLe 345
||||| ||| ||| ||| ||| ||| ||| |||
2926 GGTATTATTCTCAGCACCAGCGCTCTTACCATTTGCTTTACCGATTGCC 2975
345 uLysValIleArgPheAlaGlyThrAsnIleLeuHisMetProSerLeuA 362
::|: ||| ||| ||| ||| ||| ||| |||
2976 TAGACCTTTCAGATT.....TTAA 2995
362 snGluGluArgThrPheGlyTrpLysGlyCysValTyrLeuProAsn 378
||||: ::|: ||| ::|: ||| ::|: ||| ::|: |||
2996 ATAAGTATCATAGAGTGGTGGCGAAGAGTGGTGTGTTGATTACTGAC 3045
379 ValProLysThrAlaIleValAlaGlyAspArgThrArgLeuGlyGly 395
::|: ||| ||| ||| ||| ||| ||| |||
3046 ACT.....GACAGATTGCGTCCGTCGGTAGGGAG 3071
395 uIleLeuAlaSerValAlaAsnAlaLeuAsnGlnGluValTyrSerS 412
::|: ||| ::|: ||| ||| ::|: ||| ::|: |||
3072 TGTATTATGATAAGTTAAGAAAGCTTTGCCCGAT.....TCTT 3109
412 erValValSerSerIleThrAsnArgLeuValLeuArgAspGlnSerAla 428
::|: ||| ||| ::|: ||| ::|: ||| ::|: |||
3110 TACTCGTTCGATCAGTAGAGAAAGATTTGTAGAAATTCAAATCAAACTCTG 3159
429 LeuLeuSerHisLeuAspThrLysLeuCysAspMetPheSerGlnArgAs 445
::|: ||| ||| ||| ||| ||| ||| |||
3160 .....GTCGATACGCCAA..... 3171
445 pAlaMetIleArgGluLysProSerHisArgCysAspValPheLeuLys 462
||| ::|: ||| ::|: ||| ::|: ||| ::|: |||
3172 .....CGGCTAGACCCGGTGTTCAGTGCCCTTTTCAGATGAGTG 3211
462 roArgGluArgGluLysLeuArgGluLeuPheProGluLeuSerIleGln 478
::|: ||| ||| ::|: ||| ::|: ||| ::|: |||
3212 AGAACCAACAGGTGATGCTCCTAGAGCCCTATCTGAGTTTAATATAAAC 3261
479 PheSerAspSerValArgSerSerHisProPheAlaAsnAlaMetArgSe 495
|||: |||: |||: |||: |||: |||: |||: |||
3262 TTCTACTCATTCCGTACACTCCGATCCTCCGCTGCGTCCGGGCTCCAGAA 3311
495 rCysPheAsnGlyIlePheSerArgCysGlyAsnValCysPhePheA 512
||| ||| ::|: ||| ::|: ||| ::|: ||| ::|: |||
3312 GTTAGAATACCTCGTCGCAATATGACGAGCT...GACTATTTCGG 3358
512 spIleGlySerPheThrTyrHisValLysAlaGlyHisValAsnCys 528
::|: ||| ||| ::|: ||| ::|: ||| ::|: |||
3359 ACGTAGTGGTGGTCCCACTATTTCATTTCGTCGAGGCCACACAGGAGTT 3408
529 HisValCysAsnProValLeuAspValLysAspValLysArgArgIle.. 544
|||||: |||: |||: |||: |||: |||: |||: |||
3409 CAGGTTGACAGCGGTGTACGATGTTAAAGACGCCGCCCGTAGAGTCGT 3458
545 .....AsnGluIleLeuPheLeuSerThrAlaGlyGlyA 556
::|: ||| ||| ||| ||| ||| ||| |||
3459 TCGTCATCATCACTTGGTAAAGTAGCTTTTCGATCATCTACGCGGTCTA 3508
556 spSerTyrValSerSerAspLeuLeuThrGluAlaSerLysSerVal 572
```


1089 g...GlyLeuAlaAlaAsnAlaArgArgTyrAlaGlySerSerTyrGluTh 1105
||||| : : : ||| |||
5176 GTTGGACTTATTGAAATTCAGCCCGTAATATGGGA.....CTGGAGAC 5219
1105 rLeuSerSerLeuSerSerGlnAlaAlaGlyGlyLeuArgGlyLeuThS 1122
||||| : : : |||
5220 TTGTCTTCGATAGCTCAGCTATT.....TTTGATGTC 5244
1122 erSerThrValSerGlySerLeuValAlaArgArgGlyPheSerSer 1138
||||| : : : |||
5245TTCCTCTGGG 5253
1139 AlaValThrValThrArgAlaThrValAlaLysArgGlnValProLeuAl 1155
||||| : : : |||
5254 ATACTTACC.....TATCCTCTGGG 5273
1155 aLeuLeuSerPheSerThrSerTyrAlaIleSerGlyCysSerMetLeuG 1172
: ||||| : : : |||
5274 AGTATTAGCTTCTGCTGCTTTGTGG.....AGAGTTTTTG 5311
1172 lyleTrpAlaHisAlaLeuProArgHisLeuMetPhePheGlyLeu 1188
|| ||| : : : |||
5312 CGCCCTCGCTATGGAAGTGGGATCGAAGCTATTGCTTTAAGAAAGCTCTC 5361
1189 Gly.....ThrLeuLeuGlyAlaArgAlaSerAlaAsnThrTrpLysPh 1203
||| : : : |||||
5362 GGCAACCCGAAATTTAGGAGAGGATGCTTTTCGTAGAAGC..... 5403
1203 eGlyGlyPheSerAsnAsnTrpCysAlaValProGluValValTrpArgG 1220
5404GTAAGTATGGTGTTCAGCG 5422
1220 lylsSerValSerSerLeuLeuLeuProIleThrLeuGlyValSerLeu 1236
: ||||| : : : |||
5423 CTAAGGCTTTCGACAACT.....TCAGTC 5448
1237 IleIleArgGlyLeuLeuAsnAspThrIleProGlnLeuAlaTyrValPr 1253
: ||||| : : : |||
5449 GATGCTAAGGGTTAGTT.....AGATGTTCCGGTATTAT 5483
1253 oProValGluGlyArgAsnValTyrAspGluThrLeuArgTyrTyrArgA 1270
||| ||| : : : |||
5484 TCCTGGAGTATTGAGGATAGTGTTCATCTCTGTTTCC..... 5523
1270 spPheAspTyrAspGluGlyAlaGlyProSerGlyThrGlnHisGluAla 1286
||| ||| : : : |||
5524GATGCTGACGATGGTAGGTATAGC.....AAGCAGAA... 5559
1287 ValProGlyAspAspAsnAspGlySerThrSerSerValSerSerTyrAs 1303
||||| : : : |||
5560GTTTCCTCACTTCC 5573
1303 pValValThrAsnValArgAspValGlyIleSerThrAsnGlyGluValT 1320
: : : : :
5574 CAGATC..... 5580
1320 hrGlyGluGluThrHisSerProArgSerValGlnTyrThrVal 1336
||||| : : : |||
5581AGGTATAGTGCAGAAAGGTATCTTTCTGTTGGAG 5616
1337 GluGluGluValAlaProSerAlaAlaValAlaGluArg..... 1349
: : : : : ||| : : : : :
5617 CAGACTAAGGAAGCGTGCATCGCTTTTCATAGTTCGTCTAAGAGAAT 5666
1350GlnGlyAspProS 1354
5667 CTTAGCGTCTTACCTCGGATTCACAGATCTTTGCAAACTCGGAAC 5716
1354 erGlySerGlyThrAlaAspAlaMetAlaPheValGluSerValLysLys 1370
||||| : : : ||| : : : |||
5717 TTACTAGTGTCTAAATCCTTGTGTGGACGTGTGATTTCTATCTCTCA 5766

1371 GlyValAspAspValPheHisGlnGlnSerSerGlyGluThrAlaArgG 1387
5767 GCATCCAATGGAATGTTAAACAAG.....GCAGAAACCCCTCGCAGC 5810
1387 uValGluValAspGlyLysGlyLeuLeuProGluSerValValGlyGluA 1404
5811 AGTTAAGCGCGCGGAGTGGT...CTTAGTAACGTGTGAAGCGCATTC 5857
1404 laProThrGlnGluArgGlyArgAlaAlaAspGlyAsnThrAlaGlnThr 1420
5858 GTCCCTTCTTCCATCCCTGCTGCGGTGTTAAACCAACAGACCGAATAC 5907
1421 AlaValAsnGluClyAspArgGluProValGlnSerSerLeuValSerSe 1437
5908 TATTCGCCAGTCAGCACTCTGAGTTGGACGAGTCGCTTAGTCACAC 5957
1437 rProGlnAla.....AspIleProLysValThrG 1447
5958 ACCCGCTCTCTGGTAAACGGGAAGTGTAAACGATTTATTTCGTGCTGTT 6007
1447 InSerGluValHisAlaGlnLysGluValLysGlnGluValProLeuAla 1463
6008 CCGGCGCTGCTTAGTATATCGCAGG..... 6033
1464 ThrValSerGlyAlaThrProIleValAspGluLysProIleProSerVa 1480
6033 6033
1480 lThrThrArgGlyValLysIleLeAspLysGlyLysAlaValAlaHisV 1497
6034CGGCTGTAGAGCCTTGAG.....CGTTAATCCTACACG 6070
1497 alAlaGluLysLysGlnValGlnValGluGlnProLysGlnArgSerLeu 1513
6071 TCGGTCAATCTGTTACTCCGCATATCGGACTCTTCGTGAGGTGTTTC 6120
1514 ThrIleAsnGluClyLysAlaGlyLysGlnLeuCysMetPheArgThrCy 1530
6121 GTAAGAACCAATACTACGATCGGGAAG.....TGCAATCCTTGGCTGT 6164
1530 sSerCysGlyValGlnLeuAsp.....ValTyrAsnGluAlaThrIleA 1545
6165 CGACTGTGCCACAAAATTACTGATTTAGTTCGGATATATCCTTGGC 6214
1545 laThrArgPheSerAsnAlaPheThrPheValAspAsnLeuLysGlyArg 1561
6215 GTAATCTTCATTTGCTGCGTATTCGACAACTCCAGGGTTATCTGCGGT 6264
1562 SerAlaValPhePheSerLys...LeuGlyGluGlyTyrThr..... 1574
6265 AGTAGAGTTTCTCTTCCAGGCTTTCGGGTGATGTCACGTTGTGAAT 6314
1575TyrAsnGlyGlySerHisV 1581
6315 TCAAAATTTGTTACCGCGTGTGACGATTTTCAATGAGACCTTTTATT 6364
1581 alSerSerGlyTrpProArgAlaLeuGluAspIleLeuThrAlaIleLys 1597
6365 TGAGTTTGGCCCATCGGATCGAGACTGCCCTCGATAAATGTTGTTAGAG 6414
1598 TyrProSerValPheAspHisCysLeuValGlnLysTyrLysMetGlyG 1614
6415 TTCCTGAAGAACCTCCAGTTGTTAGTTTCGATATACACTTTC.... 6459
1614 yGlyValProPheHisAlaAspAspGluGluCysTyrProSerAspAsn 1631
6459 6459
1631 roIleLeuThrValAsnLeuValIGlyLysAlaAsnPheSerThrLysCys 1647
6460ATAGACTGT 6468
1648 ArgLys.....GlyGlyLysValMe 1654

```

|||||
6459 AGATCTAGGATTTTCCCTTTGAAGAGTTCTTTGGGGCTTTTGCATTTT 6518
|||||
1654 tValIleAsnValAlaIleSerGlyAspTyrPheLeuMetProCysGlyPheG 1671
|||||
6519 CGTCTGGAAACATCATCAAGAA.....CTTTCCGTATGATTATAGG 6559
1671 In.....ArgThrHisLeuHis 1676
::
6560 AAAATCTTCCCGTTTGCATTACGAGGTTGTGTAATAGGCGATTCCGCTGT 6609
1677 SerValAsnSerIleAspGluGlyArgIleSerLeuThrPheArgAlaTh 1693
|||||
6610 AGTTCCAATGAAGATGATGAGCGACGCTTCAAGTAATCACTGAAGCTGA 6659
1693 rArgArgValPheGlyValGlyArgMetLeuGlnLeuAlaGlyGlyValS 1710
|||||
6660 AAACCGA.....ATTGGTTTTCGGAGAGGTTCTCCCGGAAATGG 6700
1710 erAspGluLysSer.....ProGlyValProAsnGln 1720
::
6701 CACGGTATCGTCCGTCCTCTCGAGATTCCCGTCCCGGGTCCGCTAACAGT 6750
1721 Gln.....ProGlnSerGlnGly 1726
|||||
6751 CCTTTGTTTCCACCGACATACGGTGAGATTGAGGAACTAGTGACGGTTC 6800
1727 .....AlaThrArgThrIleThrProLysSerGlyGlyLysAlaLeuS 1741
|||||
6801 TTCGTATAGTCCCGAGTGATAGGTGGACACACGGTTCACACTCAATGT 6850
1741 erGlu..... 1742
||
6851 CCACTTGAAGGGGACCGGAACCTAGATTCCGATCGGTGATAGCGATGATPCG 6900
1743 .....GlySerG1 1745
6901 ATACCTTTGGATCGTGACTTTTGACGAGATCGGAGTCTTGGAGGAAGCGG 6950
1745 y..... 1745
6951 CACGCTCGACGCTTTCGCTAGGGTTGAGGTTGCGTGTGGAGATGTCCG 7000
1746 .....ArgGluValLys..... 1749
7001 GTAATTTATTAACTTAAGACCGCTTAAATTTCTTATGGTTAGCTCTT 7050
1750 .....GlyArgSerThr..... 1753
7051 CAGTGATATTTAACGGCAGATCATCGACGGCCAGTTTATATCGCACTGTT 7100
1754 .TyrSerIle..... 1756
:::
7101 ATTCAGTTTCGGCGCTGTTTACCGTTTGAGTTCTTCTCAGGCACGTTTGTG 7150
1757 .....Trp..... 1757
7151 CTGTGAGGTGTTGTCAAGAGCTTGGCATTTACAGTTCGTAAACGCTAGT 7200
1758 ...CysGluGlnAspTyrValArgLysCysGluTrpLeuArgAlaAspAs 1773
7201 ATCTCGCACGAAGAACCGTT.....GAATGGCTTTTCAGSTTTTTC 7241
1773 nProValMetAlaLeuGluProAspTyrThrProMetThrPheGluValV 1790
7242 CGCATTTTGTG...TAGCATTCGCTTATTTATCCCTCGTGTTCGGAGT 7288
1790 alLysThrGlyThrSerGluAspAlaValGluTyrLeuLysTyrLeu 1806
7289 TG.....CAGAGGGTTCGAATTTTACTTAAGTTACCTG 7320
1807 AlaIleGlyIleGluArgThrTyr..... 1814
:::

```

```

7321 TTCCTGTTGCCGACAGAGGTTTTCACGCATAAACCCGAGGTTCAAGTTTG 7370
1815 ...ArgAla...LeuLeuMetAlaArgAsnIleAlaValThrThrAlaG 1829
|||||
7371 CGTTAGAGCTCCAGTTCGTGTGGTTTAGGAACGTGGGACTT.....GTAG 7414
1829 luGlyValLeuLysValProAsnGlnValTyrGluSerLeuProGlyPhe 1845
|||||
7415 AAGATGGTAGGAAGTCTTTATCCGAAGTGTGAACCTCTTGGACTCCTTG 7464
1846 HisVal...TyrLysSerGlyThrAspLeuIlePheHisSer..... 1858
|||||
7465 AAGGTGTCTAACCGTAAAGGGAAGTCAGCGGTTATTGAAAGTGATAGTTC 7514
1858 ..... 1858
7515 CGATTCTGAGACCGGTAGTCTATCCGGTTCATCTCTCTCGGATCGATTA 7564
1859 .....ThrGlnAspGlyLeuArgVal..... 1865
7565 ATGAGGAAGTTTAAAAACGGCGAAGCTAGCGTGTGTATCTGAAAAAGCT 7614
1865 ..... 1865
7615 TCGTTGATGAAGCATGCGCGCCGCACTAAGCGGTTGGAACCAAGATCTTGG 7664
1866 .....ArgAspLeuProTyrValLeuIleAlaGluLysGlyIlePheT 1880
|||||
7665 TGGTAAGAAAGACTTCTCTATTATT.....AGGAGAAATTTC 7702
1880 hrLysGlyLys..... 1883
7703 CCGCGGTGAGTCTTCAAGTACAAAAATCGTAGGAGCTTGTGCGAGTAC 7752
1883 ..... 1883
7753 TTGACGACCTTAAATTCCTGTTCTGTTGCGTGGCGGACTTGTATCCCGT 7802
1883 ..... 1883
7803 CGCAGGTGACATAAGTACAGAGATGACCAACCCATGCGTGAATTC 7852
1883 ..... 1883
7853 ATTATTCGAAAAAGTAACGTTGTATGAATTCATGAAAAATTAAGCTCG 7902
1883 ..... 1883
7903 TACTGGGACGAGTTAAAGACGCGTGGTTTCGACAGGAATTTGGCGAAGAT 7952
1884 .AspValAspAlaValValAlaLeuGlyAsp...AsnLeuPheVal.... 1897
|||||
7953 GGAITTTGGACGATGACGTTGGTGTGTCGATTTTAACCTTGAAGGTGATAG 8002
1897 ..... 1897
8003 TGGGAAGGCATGTTACTCGTCTCTTCAGAAATACTGTGAATTCACAGAG 8052
1898 .....CysAspAspIleuVal..... 1903
8053 TTCATGTTTTCGACGACGCGTGTGGTGTTCGCCGGTTCGAAAAAACG 8102
1904 .....PheHisAspAlaIleAsnLeuIleGlyAlaLeuLysV 1916
8103 CGATTTCCCTTTTGATCTCTAGTCAACCCCAATTCGTCTCGGAACICAT 8152
1916 alAlaArgCysGlyMetValGlyGlu.....SerPheLysSerPheGlu 1930
|||||
8153 TTCTACGT...GCCGTGGACGCAAAAGATTTGACGTTCACTAATTCGGAG 8199
1931 TyrLys.....CysTyrAsnAlaProGlyGlyGlyLysThrTh 1945
8200 CATAGTCTTATAGTTTACGAGGCTCCTCCCGTGTGGGAAGACGCAATTC 8249

```



```

1755 eriletrpCysGluGlnAspTyrValArgLysCysGluTrpLeuArgAla 1771
      |||||
7738 .....AGCTTGGCGAGTACTTGACGACC 7761

1772 AspAsnProValMetAlaLeuGluProAspTyrThrProMetThrPheG1 1788
      |||
7762 TTAATTCCTCGTTCGGTGGCGGACTTGTATCCGCGCGACGTGA 7811

1788 uValValLysThrGlyThrSerGluAspAlaValValGluTyrLeuLysT 1805
      |||||
7812 CATACG.....TACAGAAG. 7827

1805 yrLeuAlaIleGlyIleGluArgThrTyrArgAlaLeuMetAlaArg 1821
      |||
7828 .....ATGACCAACGCCATCGTCAATTCTATTATTCGCAA 7863

1822 AsnIleAlaValThrThrIleGluGlyValLeuLysValProAsnGlnVa 1838
      |||||
7864 AAGTAACGTTGTATGAACCTTCATGGAATA.....AGTCGTA 7904

1838 lTyrGluSerLeu.....ProGlyPheHisValTyrLysSerGlyThrA 1853
      |||||
7905 CTGGGACGAGTTAAGACCGCGGGTTTCACAGGAAATTGGCGAAGATGG 7954

1853 splLeuIlePheHisThrThrGlnAspGlyLeuArgValArgAspLeuPro 1869
      |||||
7955 AATTG.....GACGATGACGTGGTGTTCGATTTTAAC 7989

1870 Tyr...ValLeuIleAlaLysGlyIle.....PheThrLysGlyLy 1883
      |||||
7990 TTGAAGGTATAGTGGGAAGGATGCTACTCTCTTCAGAAATACTGT 8039

1883 sAspValAspAlaValAlaLeuGlyAspAsnLeuPheValCysAsp. 1899
      |||||
8040 GAATCTCAGAGTCTATTTTGGACACGCGTTTGGCTTGGCCCTG 8089

1900 .....AspIleLeuValPheHisAspAlaIleLeuIle 1911
      |||||
8090 GTCCGAAAAACCGGATTTTCGTTTGTATCTAGTCAACGCAATTCGTT 8139

1912 GlyAlaLeuLysValAlaArgCysGlyMetValGlyGlu.....SerPh 1926
      |||||
8140 GTGCGCACTCTTCTACGT...GCCGTGGACGCAAGATTTCAGCTT 8186

1926 eLysSerPheGluTyrLys.....CystyrAsnAlaProGlyGlyG 1941
      |||||
8187 CACTAATTCGGAGCATAGTCTTATAGTTTATGAGGCTCTCCCGGTGGT 8236

1941 lLysThrThrThrLeuValAspGluPheValLys.....SerProAsn 1955
      |||||
8237 GCAAGACGCATTCCTTGGTAACGTATTCGCGGATTTATGTACTAAAGTG 8286

1956 SerThrAlaThrIleThrAlaAsnValGlySerSerGluAspIleAsnMe 1972
      |||||
8287 TCTTCCTAGTGTGACCGCTAACAGAATTCCGCAAGCGGAGATTCTCA 8336

1972 tAlaValLysLysArgAspProAsn.....L 1981
      |||||
8337 AAGGATCTCTAGAGAGTTAATCGATAGAAAAATGCTGTCAACATGTGA 8386

1981 euGluGlyLeuAsnSerAlaThrThrValAsnSerArgValValAsnPhe 1997
      |||||
8387 TTAACACCGCGGTAGAGTGTTCACCGTGACTCTATTGTATGACCAAT 8436

1998 lLeValArgGlyMetTyrLysArgValLeuValAspGluValHisMetMe 2014
      |||||
8437 ATT...CGACTTAAGACTGACTGTGTGTGTTCGATGAGCTGTTTATGCT 8483

2014 thisGlnGlyLeuLeuGlnLeuGlyValPheAlaThrGlyAlaSerGluG 2031
      |||||
8484 TCATCTGGAGCAATCGGACGAGTGTGTGAATTCATCTCATGAAGCGCG 8533

```

```

2031 lyLeuPhePheGlyAspIleAsnGlnIleProPheIleAsnArgGluLys 2047
      |||||
8534 TGGTTTTCITCGGTGATAGTACAAATTCATTACATTCACCGGAACGAT 8583

2048 Val.....PheArgMetAspCysAlaValPheValProLysLysG1 2061
      |||
8584 CTGTGGTGTTCGTGTTTACATGACATAGACGCATTTATA...GAACACA 8630

2061 uSerValValTyrThrSerLysSerTyrArgCysProLeuAspValCysT 2078
      |||||
8631 GCACCCATTTACGGTGAGGTGCTTACAGGTGCTCTGGGACATCTGTG 8680

2078 yrLeuLeuSerSerMetThrValArgGlyThrGluLysCysTyrProGlu 2094
      |||||
8681 AGTGGTGTCT.....AAATTCATCCGAGA 8706

2095 LysValValSerGlyLysAspLysProValValArg...SerLeuSerLy 2110
      |||||
8707 CATGTAGCTACTGCTAACACATGGAAGTGTGGGAAATCGAGTGTCTAT 8756

2110 sArgProIleGlyThrThrAspAspValAlaGluIleAspAlaAspValT 2127
      |||||
8757 AGAGTAAATAATGGATGTGACGATGTCTTACGACAGTAGTGCGAAGT 8806

2127 yrLeuCysMetThrGlnLeuGluLysSerAspMetLysArgSerLeu... 2142
      |||||
8807 ATATCGTTTACACACAAAGCAGAAAGCAAGCACTGCAGAAAGCACCTTGGT 8856

2143 .....LysGlyLysGlyLysGluThrProValMet...ThrValHI 2155
      |||||
8857 CGTTTAACCGTAGCGCGAACTAAAGCGGTGCTGTGTGAACACAGTACA 8906

2155 sGluAlaGlnGlyLysThrPheSerAspValValLeuPheArgThrLysL 2172
      |||||
8907 TGAAGTTCAAGGGGAAACGTATAAAGGTCCTGCTTGGTAAGGTGCAAT 8956

2172 ysAlaAspAspSerLeuPheThrLysGlnProHisIleLeuValGlyLeu 2188
      |||||
8957 ACCAAGAGGACACTCTTCTTCTGTCGATAATCATCTGCTGTGGCCTTA 9006

2189 SerArgHisThrArgSerLeuValTyrAlaLeuSerSerLysLysLeuAs 2205
      |||||
9007 ACCAGCATGTGACACTCTTACTTACTCTGTTTGAATAGTAGAGATA 9056

2205 pAspLysValGlyThrTyrIleSerAspAla 2215
      |||||
9057 CGACAGACTGCTTCTAGTATAGATGAAGCG 9087

seq_name: gb_vil:CTAF001623

seq_documentation_block:
LOCUS CTAF001623 19249 bp RNA VRL 15-NOV-1999
DEFINITION Citrus tristeza virus, complete genome.
ACCESSION AF001623
VERSION AF001623.1 GI:2098825
KEYWORDS
SOURCE
ORGANISM Citrus tristeza virus.
Viruses; ssRNA positive-strand viruses, no DNA stage;
Closteroviridae; Closterovirus.
REFERENCE
1 (bases 1 to 19249)
AUTHORS Yang,Z.N., Mathews,D.M., Dodds,J.A. and Mirkov,T.E.
TITLE Molecular characterization of an isolate of citrus tristeza virus
that causes severe symptoms in sweet orange
JOURNAL Virus Genes 19 (2), 131-142 (1999)
MEDLINE 2007030
PUBMED 10541017
REFERENCE
2 (bases 1 to 19249)
AUTHORS Yang,Z.N., Mathews,D.W., Williams,A., Dodds,J.A. and Mirkov,T.E.
TITLE Direct Submission
JOURNAL Submitted (29-APR-1997) Department of Plant Pathology and
Microbiology, Texas A&M Agricultural Experiment Station, 2415 E.
Hwy 83, Weslaco, TX 78596, USA

```


696 euThrMetCysArgAlaGlnIlysSerProGly.....ThrGlu 708
 ::: :::::::::::::: |||
3901 ATTGTATACCAAAATCGCGGGTTTCTCCGAGAATCAGCGAGATTAAACGG 3950
 ::: :::::::::::::: |||
709 valThrTyArqSerLeu.....ValProSe 717
 ::: |||||
3951 TTACGCATATCGTCGGTGTCATAGTGACTTGTATTAGAATTAAGAAITTCCTCG 4000
 ::: :::::::::::::: |||||
717 rPheValGlyLysSerLeuValPheIleproValValAlaGlySerServ 734
 ::: :::::::::::::: |||
4001 GTACTCTAGTAAGACTCGTAGTGTTTACA..... 4031
 ::: :::::::::::::: |||||
734 alserPheIylsThrIleValLeuAspSerAspPheValAspArgIleTy 750
 ::: :::::::::::::: |||||
4032 ..GGTGCCTATTCTGTATTGGACGCTAAGTTTGTGAGTCGTGTATC 4079
 ::: :::::::::::::: |||||
751 SerTyArAlaLeuAsnThrIleClyThrPheGluAsnArgThrPheGluTy 767
 ::: :::::::::::::: |||
4080 GAATACGTGGTTTAATAATGTGTGTGTGAATCTAAGACTTTTGAATG 4129
 ::: :::::::::::::: |||
767 rAlaValGlyAlaValArqSerGlnLysThrHisValIleThrGlySerA 784
 ::: :::::::::::::: |||||
4130 GACTTGGAACTACATCAAGTCTGTAACTCAGAGTGGTCATCATCGTGGGA 4179
 ::: :::::::::::::: |||
784 rgValValHisSerLysValAspIleSerProAspSPmetTrpGlyLeu 800
 ::: :::::::::::::: |||
4180 AGATCATCCATAGGACGCTTCTATCGCATTTGGAGTATTTCGATGGCTTT 4229
 ::: :::::::::::::: |||
801 valValAlaValMetAlaGlnAlaIleLysAspArgAlaySerIleAr 817
 ::: :::::::::::::: |||
4230 TCCGCGATTATGCTCTCCGCGGGCTGAAGAGTAGG...CAGAACGCAGCA 4276
 ::: :::::::::::::: |||
817 gSerTyArSnPheIleLysAlaSerGluGlyLysLeuAlaGlyValPheIL 834
 ::: :::::::::::::: |||
4277 GCGTTTTCTAGCGCTAGCGGCTTTTCAGGAGACAMTTCTCTTTTCG 4326
 ::: :::::::::::::: |||
834 ysLeuPhe...PheGlnThrValGlyAspCys..... 843
 ::: |||
4327 AATTGTGTAGGTTGCCATTTTCAGAGAAGTCTCGTACTATTTCCTGGCG 4376
 ::: :::::::::::::: |||
844 PheSerAsnAlaValASerValTyralalysAlaMetValHisAspAsnph 860
 ::: :::::::::::::: |||
4377 ATTCAGGAGGCTGTGCGAGAGGTGCGTGAATCATCATCTCAGAAATCGCTT 4426
 ::: :::::::::::::: |||
860 eAsnValLeuGluThrLeuMetSerMetProArqAlapheIleargLysV 877
 ::: :::::::::::::: |||
4427 TAATATGACTTTCGTAGACTTCTCTGATCT.. 4457
 ::: :::::::::::::: |||
877 alProGlySerValValValThrIleCysThrSerGlyAlaSerAsp... 892
 ::: :::::::::::::: |||
4458CTGCTCACYATCAGTGAGTACTCRGAGTTAGATGTT 4493
 ::: :::::::::::::: |||
893 ArgLeuGluLeuArgGlyAlaPheAspIleSerLysGluThrPheGlyAr 909
 ::: :::::::::::::: |||
4494 CCTATTGATTTACC GGTTTCGGTGCATACCGAGGAGGACGAGGTAA 4543
 ::: :::::::::::::: |||
909 glylsLeuLysAsnSerArqLeuArqValPheSerArqAlaIleValGluA 926
 ::: :::::::::::::: |||
4544 A.....CTTATAAAG 4554
 ::: :::::::::::::: |||
926 spSerIleLysValMetLysAlaMetLysThrGluAspGlyLysProLeu 942
 ::: |||||
4555 ATGGAGTTAAGTCGGCTTTACTGCCGAANGCA..... 4586
 ::: :::::::::::::: |||
943 ProIleThrGluAspServAlTyralaPheIleMetGlyAsnValSerAs 959
 ::: :::::::::::::: |||
4587 ...GTTTCGTGACGAAGCAACTCATTTCTCAAAGGGCGGCACGTCTCAGA 4633
 ::: :::::::::::::: |||
959 n...ValHIScysThrArqAlaGlyLeuLeuGly..... 969
 ::: |||
4634 GATCGTCCCTGTGGGGATGAAGAAGTTTATAGCGCGGCGGTCTTCACA 4683

970GlySerLysAlaThrValValSerSerValSerLysGlyLeu 983
 4684 GTTAAATYGGTCCGGTAGTATCGCCTCTCTAATGCTAGTTCTGGTGT 4733
 984 ValAlaArgGlyAlaAlaThrLysAlaPhe 993
 4734 TTGGGGCCGGAGCTGCCCTTCACTGTTTACAGTTGTTGAAGAAGTT 4783
 993 993
 4784 GGGTGGGGTTTTTGCATGGCTAAAGAGTTGTTGCACITGTTGCGGAC 4833
 994SerGlyIleThr 997
 4834 GATTAGTTGCTTCGTTATCGCGCCTAAAGTTCGGTACAACTCATTTAAC 4883
 998 SerPhePheSerThrGlySerLeuPheTyrAspArgGlyLeuThrClnuAs 1014
 4884 TCACTCTTACATGTTACGTATTAAAGCTTTCGCGCTAATTTCTCGACGA 4933
 1014 pGluArgLeuAspAlaLeuValArgThrG 1024
 4934 CTCGATTTTACCGCTGTTTGGGACACGGTAATGAACITTAACTAGNATCG 4983
 1024 luAsnAlaIleAsnSerProValGlyIleLeuGluThrSerArgValAla 1040
 4984 TCAATGGGTGACCAACGCAAGTTGGAGTGTTTTCGCAACGTC 5027
 1041ValSerLysValValAlaGlyThrLys 1049
 5028 CTTTCGGTTAATTCCTAAAGCTGTGTTCTGTTGTAAGGACCTCGCAAGA 5077
 1050GluPheTrp 1053
 5078 CTTTCGGACATGTGAGTTAGTTTTCGCTTTTCTACGGAGTCTTCA 5127
 1053 er 1053
 5128 GTAAAGACTACGCTAAACTATGCGTTTAGTTTGAAATATTCGATATC 5177
 1054GluValSerLeuAsnAspPheThrThrPheVal 1064
 5178 GGACTCAATGAAATTTTCAGCCGCTAATATGGATTTAGACACTTGTTTC 5227
 1064 lLeuArgAsnLysValLeuIleGlyIlePheValAlaSerLeuGlyAla 1081
 5228 GATGGCTTCGCATATTTTGTATGTATTACTTACCTATCTCTGGGAGTA 5277
 1081 laProIleAlaTrpLysTyrArgArgGlyIleAlaAlaAsnAlaArgArg 1097
 5278 TTAGCTTTTCCTATCGTTTCGCGAGGGTTTTCGGCGCCCTCGCTATGGA 5327
 1098 TyrAlaGlySerSerTyr 1103
 5328 ...GTAGGATCGAACTATGTTTAAAGAAAACCTTTCGCAACCTCAAT 5374
 1104GluThrLeuSerSerLeuSerGln 1113
 5375 TTTAGGAGAGATGCTTTCCTAGACGGTAGTATGTTGTTGACGGCTA 5424
 1113 laLa 1114
 5425 AGGCTTTCGACAAACTTTTCGCTCGACGGTAAAGGTTTAGTTAGGTTC 5474
 1115GlyGlyLeuArgGlyLeuThrSerSerThrValSerG 1127
 5475 GGTGTTATACCTGGAATACTGAGGATAGTGTCTCTCTCTCTCTCTCCGA 5524
 1127 yGlySerLeuValValArgArgGlyPheSerSerAlaVal 1141
 5525 TGTCGCGCATGGGTAGGTATATGAACACGAAAGTTTCTCTCGCTCCCA 5574
 1141 hrValThrArgAlaThrValAlaLysArgGlnValProLeuAlaLeuLeu 1157

5575 ...||| ... ||||| |||
TGATTACGTATAGY.....CCCAAAGG.....TTG 5600
1158 SerPheSerThrSerTyrAlaIleSerGlyCys..... 1168
||| ||| ... ||| ...
5601 TCTTTTCGTTCGAGCAGGCTAAGGAAGCGTCATCGCTTTCATAGTCGA 5650
1169SerMetLeuGlyIleTrpAlaHisAlaLeuP 1179
... ||| |||
5651 TTCGCTAAGAAGATTTAGCATCTTGACCTCAGAGATTACGAATCTT 5700
1179 roArgHisLeuMetPhePheGlyLeuGlyThrLeuLeuGlyAlaArg 1195
: : : : : ||| : : : : : |||
5701 TCCAAACACTCGGAATTAACAGTAGTGGTGTAAATCTTGATGGACATGTG 5750
1196 AlaSerAlaAsnThrTrpLysPheGlyPheSerAsnAsnTrpCysAl 1212
: : : : : ||| : : : : : |||
5751 AGTTCTTACTCTACGCGTCCAAATGCTGTAGTTAACAGGCGGAACGC 5800
1212 aValProGluValValTrpArgGlyLysSerValSerSerLeuLeuLeuP 1229
||| ||| ||| : : : : : |||
5801 CGTTCAGCGGTAAAGCAGCGGGAGTGGTCTTAGCAACGCTGTGAGGC 5850
1229 roIleThrLeuGlyValSerLeuIleIleArgGlyLeuLeuAsnAspThr 1245
||| : : : |||
5851 GCATTCGGCCTTCCTTTCGCCCT..... 5873
1246 IleProGlnLeuAlaTyrValProValGluGlyArgAsnValTyrAs 1262
: : : : : ||| : : : : : |||
5874GCTCGGTGTGTTAAATGA 5890
1262 pGluThrLeuArgTyrTyrArgAspPheAspTyrAspGlu..... 1275
: : : : : ||| : : : : : |||
5891 ACAGACCAGCGAATACTACTCTGCTAGTAGCAGAGTCTGAGTTGGATGAGT 5940
1276GlyAlaGlyProSerGlyThr 1282
||| ||| ||| : : : : : |||
5941 CATTGCTTAGTGACACACCGGCTCTTCGTGGTAACGGGAAGTCTAGCGAT 5990
1283 GlnHisGluAlaValProGlyAspAsnAspGlySerThrSerSerVa 1299
: : : : : ||| : : : : : |||
5991 TTGTTTCGTGCTGCTTCGGA.....CGTGTCTCAAT 6022
1299 lSerSerTyrAspValValThrAsnValArg.....AspValG 1312
: : : : : ||| : : : : : |||
6023 ATGTTGTAGGCAGCGCTGTGGAAAGCTTGAGGCGTTAGCCCTACACGTCG 6072
1312 ly..... 1313
||| : : : : : |||
6073 GTGACTACTGTGACTCGTGCAATTATGGACTCTTTTCTGAGGTTGTTCGTA 6122
1314 SerThrAsnGlyValThrGlyGluGluThrHisSerPro..... 1328
||| ||| : : : : : ||| : : : : : |||
6123 AGAACCAACATCGCAGTC...GGGAAGTGTGAATCCTTCGCGCGTCGATG 6169
1329ArgSerValGlnTyrThrTyrValGluGluValAla 1341
: : : : : ||| : : : : : |||
6170 TGCCACTAAGATTCTGATTGGTAGTCGATATATTCTTGGCGCAATC 6219
1341 laProSerAlaAlaValAlaGluArgGlnGlyAspProSerGlySerGly 1357
||| : : : : : ||| : : : : : |||
6220 TTCAATTCGGTCCGTATTCCACAACATCCAGGATTATCCGCGCGCAGTAGA 6269
1358 ThrAla.....AspAlaMetAlaPheValGluSe 1367
: : : : : ||| : : : : : |||
6270 AATTCTCTTCAGGTTCTTGACGCGGCTGTGATGTGTAAATTCAAAA 6319
1367 rValLysLysGlyValAspValPheHisGlnGlnSerSerGlyGluT 1384
: : : : : ||| : : : : : |||
6320 TCTTGGTACCGGTGGCAGACGTTTTCAATGAGACCTTTTACTCRAGCC 6369
1384 hrAlaArgGluValGluLeuAspGlyLysGlyLeuProGlu..... 1398
||| : : : : : ||| : : : : : |||

6370 TGGCCCATCGGATGAGGACTGCTTCGATAAATGTTGTGTAGAGTTCCTG 6419
1398 1398
6420 AAGAACCCTCCCGTTTGTAGTTTCGATATACACTYTCATAGAGTGTAG 6469
1398 1398
6470 ATCTAGATTTCCTTTGAAGAGTTCCCTTTGGGGCTTTTCCGTTTGTG 6519
1399SerValValGlyGluAlaPro 1405
6520 TTTGGAACATTGTTAAGRAACTTTCATATGCATTTACGGAAGTCTCCCC 6569
1406ThrGlnGluArgGlyArgAlaAlaAspGlyAsnThrAlaGl 1419
||| : : : : : ||| : : : : : |||
6570 GTTTCGGTTACGAGGTTGTGTAATAGG.....GTGTTCCG 6604
1419 nThrAlaValAsnGluGlyAspArgGluProValGlnSerSerLeuValS 1436
: : : : : ||| : : : : : |||
6605 TCGCAGTTCACAGGAAGGTGACGTGCTAGCTTCNA..... 6641
1436 erSerProGlnAlaAspIleProLysValThrGlnSerGlu.....Val 1450
: : : : : ||| : : : : : |||
6642GTAATCACGGAAGCCGAAAAACCGAATC 6668
1451 HisAlaGlnLysGluValLysGlnGluValProLeuAlaThrValSerGl 1467
: : : : : ||| : : : : : |||
6669 GGTGTTTTCGAGAAAGTTCTCCCGAGATGGCACGTTATCGTGTCTCT 6718
1467 yAlaThrProIle..... 1471
||| : : : : : |||
6719 CGAGATACCCGTACCGGGGTCGCTGATRGTCCTTTGTTTCCACCGACGT 6768
1472ValAspGluLysProAlaProSerSerValThrThrArgGlyVal 1485
: : : : : ||| : : : : : |||
6769 ACGGTGAGATTGAGGATTAGTAACGATTCTCTGACAGTCCCGCAGTG 6818
1486 LysIleIleAspLysGlyLysAlaValAlaHisValAlaGluLysLysGl 1502
: : : : : ||| : : : : : |||
6819 ATAGGTGTGGACAATGTTTCAGATTTCAGTCCGCTGAAAGCGGAAGCGGA 6868
1502 nValGlnValGluGlnProLysGlnArgSerLeuThrIleAsnGlu..... 1517
: : : : : ||| : : : : : |||
6869 ACTAGATTCCGGATGGTATAGCATGATTCATACCTTTGGATCGCGACT 6918
1518GlyLysAlaGlyLysGlnLeuCysMetPhe 1527
||| : : : : : |||
6919 TTGAGCAAAATCGGAGGTCTTTCGCGGAAGCGGT..... 6950
1528 ArgThrCysSerCysGlyValGlnLeuAspValTyrAsnGluAlaThrIl 1544
: : : : : ||| : : : : : |||
6951AACGTTGACGTCTTTCGT.....TT 6970
1544 eAlaThrArgPhe.....SerAsnAlaPheThrPheValA 1556
: : : : : ||| : : : : : |||
6971 AGGGCTGAGGTTCCGTGGAAAATGTTGGTAATTCGTTT..... 7010
1556 spAsnLeuLysGlyArgSerAlaValPhePheSerLysLeuGlyGluGly 1572
||| : : : : : ||| : : : : : |||
7011 ..AACTTA.....AGAGCGGTAAATTTTCTATGGGTTAGCTCTTCAG 7052
1573 TyrThrTyrAsnGlyGlySer..... 1579
: : : : : ||| : : : : : |||
7053 TGTRTATTTAACGCGGTTCAATGGCCAGTTTACGTCGCGACTGTTATT 7102
1580HisValSerSer..... 1583
||| : : : : : |||
7103 CAGTTTCGCGCAACTGTTTACCGTTTGGTTCCTTTAGACACGTTTGTAGTG 7152
1584 ..GlyTrpProArgAlaLeuGluAspIleLeuThrAlaIleLysTyrPro 1599
||| : : : : : ||| : : : : : |||
7153 TGAGGTGG.....CTGTCAAAGAGCTTAGCAGCTAGCTGTTGCTAACGCT 7196

1600 SerValPheAspHisCysLeuValGlnLys.....TyLysMetClyGlyVal 1616
|||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
7197 AGTATTGTGCAGCAAAACGGTTTCACAGCTTTCTTCAGCTTTTCGCCGATT 7246

1610TyrLysMetClyGlyVal 1616
::: ::: ::: ::: ::: ::: ::: ::: ::: :::
7247 TTTGTACGATTCGCGTTACTTATCCGACGCTGTTTTCTTCGCTGTTCGAGRGG 7296

1616 alProPheHisAlaAsp.....AspGluGluCysTyrPro 1627
||| ||| ||| ::: ::: ::: ::: ||| ||| ||| ||| ||| ||| ||| |||
7297 TCGATTTTTTACTTAAGTTATCTATCTCGTTGCCAAGAGGTGCTTACG 7346

1628 SerAspAsn.....ProIleLeuThrValas 1636
::: ::: ::: ::: ::: ::: ::: ::: ::: :::
7347 CGTAACCGTAGTGTCGAGTTTGGCTAGAGCTCCAGTTCGTGTGATTAG 7396

1636 nLeuValGlyLysAlaAsnPheSerThrLysCys..... 1647
::: ||||| ::: ::: ::: ||| ::: ::: ::: ::: ::: :::
7397 GGACGTGGCACTTGTAGAAGACGGCAAGAGCTTTAACCCAAGTGTGA 7446

1648ArgLysGlyGlyLysValMet 1654
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7447 ACTCTTTGGACTCATTTGAAGTGTCCGACCCTGAAGGAAGTCCAGCACT 7496

1655 ValIleAsnValAlaSerGlyAspTyrPheLeuMetProCysGlyPheGI 1671
::: ::: ::: ::: ::: ::: ::: ::: ::: :::
7497 ATT..... 7499

1671 nArgThrHisLeuHisSerValAsnSerIleAspGlu...GlyArgIle. 1686
::: ||| ||||| ||||| ::: ||| ||||| ||||| ||||| ||||| |||||
7500GAAAGTGATAATTCCGACTCCGACGAGCGGTCTGTCGT 7536

1687SerLeuThrPheargAlathrArgArgvalPheGlyValGlyArg 1701
||||| ::: ||| ::: ::: ::: ::: ::: ::: ::: ::: :::
7537 TCGGTTCCTATTTTCATTCGGATCGGTTAATGAG.....GAA 7571

1702 MetLeuGln...LeuAlaGlyClyValSerAspGluLysSerProGlyVa 1717
::: ||| ::: ||||| ||||| ::: ||||| ||||| ||||| ||||| ||||| |||||
7572 GNTTTAGAAGCAGTCAACGGTGGTGTGTATCCCAAAAAGCTCCA..... 7616

1717 lProasnGlnlProGlnSerGlnGlyAlaThrArgThrIleThrProL 1734
::: ::: ::: ::: ::: ::: ||| ::: ||||| ||||| ||||| ||||| |||||
7617CTGATGAAGCATGCGGCTGCACTAAGCAGGTGTGGACGCAGATC 7659

1734 ysSerGlyClyLysAlaLeu.....Ser 1741
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7660 TTATGGTGAATAAGGATCTCCTATTGTACGAGAAGTTTACC GCCGT 7709

1742 GluGlySerGlyArgGluValLysGlyArgSerThrTyrSerIleTrpCy 1758
|||::: ||||| ::: ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7710 GAATCTTCAGGTACGAACCCGCCAGGA..... 7736

1758 sGluInAspTyrValArgLysCysGluTrpLeuArgAlaAspAsnProv 1775
|||::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7737AGCTTGTGCGTGTACTTAACGACGTTGAATTTCTT 7770

1775 aMetalalaLeuProAspTyrThrPrometThrPheGluValValLys 1791
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7771 CATTCGGTTCGGTGCCTACCTGTATCCGTCGACGCTGCATCAACG... 7817

1792 ThrGlyThrSerGluAspAlaValValGluTrpLeuLysTyrLeuAlaIl 1808
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7818TACAAGAG..... 7826

1808 eGlyIleGluArgThrTyrArgAlaLeuLeuMetalaaArgAsnIleAlay 1825
::: ::: ::: ||| ::: ::: ::: ::: ::: ::: ::: ::: :::
7827 ...ATGACCAACGCCATGCGCGAGTTTATTACTCCAAAAAGGTAACT 7872

1825 alThrThralaGlyValLeuLysValProasnGlnValTyrGluSer 1841
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7873 TCACGAACTTCACGGAAATTTG.....ACCTCGTACTCGGATGAA 7913

1842	LeuPro.....GlyPheHisValTyrLysSerGlyThrAspLeuLeuLeph	1855
7914	TTAAAGACGGTATGGTTTGATAGGAAGTTGCGAAGATGGATTTA.....	7958
1856	eHisSerThrGlnAspGlyLeuArgValArgAspLeuProTyr...Vall	1872
7959GACGATGACGTTAGTGTCTCGATTTTAACTTGAAGGTAA	7998
1872	eulleAlaGluLysGlyIle.....PheThrLysGlyLysAspValAsp	1886
7999	TAGTGGGAGGCATGTCACCGCTCTTTCAGAAATACTGTGAACCTCTCAC	8048
1887	AlaValValAlaLeuGlyAspAsnLeuPheValCysAsp.....	1899
8049	GAGTTCATCTTTTGTATGACGAGTGTGTCTTCCCGCGGTCCGAAAAA	8098
1900AspIleLeuValPheHisAspAlaIleAsnLeuIleGlyAlaLeuL	1915
8099	GGGTGACITTTGCTGTATCTCTAGTCAACCAATTTGTTGCTCGCAACT	8148
1915	ysValAlaArgCysGlyMetValGlyGlu.....SerPheLysSerPhe	1929
8149	CTTCTCTGGT...GCCGTGGACGGTAAGATTTCAGCTTCACATAATTCG	8195
1930	GluTyrLys.....CysTyrAsnAlaProGlyGlyGlyLysThrTh	1944
8196	GAGCATAGTCTCATGTTTATGAGGCTCTCCCGGTGTGGAAAGACGCA	8245
1944	rThrLeuValAspGluPheValLys.....SerProAsnSerThrAlat	1959
8246	TTCTTTGGTAAAGTATTTCCGCCGATTAFTGTPACCAAGTGTCTGTCTGG	8295
1959	hrIleThrAlaAsnValGlySerSerGluAspIleAsnMetAlaValLys	1975
8296	TGGTGACCGCTAACAAAAATTCGACAGCGGAGATTTCTCAAGNATCTCT	8345
1976	LysArgAspProAsnLeuGluGlyLeu.....	1984
8346	AGGGAGTTAATGGATGAGAAGATGCTGCTAAACACAGTTGTTAAACCCG	8395
1985	.AsnSerAlaThrThrValAsnSerArgValValAsnPheIleValArg	2001
8396	CGGTAGAGTGTTCATCTGCGACTCTTATCTGATGAACCATATT...CGAC	8442
2001	lyMetTyrLysArgValLeuValAlaAspGluValHisMetHisGlnGly	2017
8443	TCAAGACTATTGTGTGTTGTCGACGAGTGTTTATGGTTCATGCTGGA	8492
2018	LeuLeuGlnLeuGlyValPheAlaThrGlyAlaSerGluGlyLeuPhePh	2034
8493	GCAATTTGGAGCGGTTGTGTGTTGTCGACGAGTGTTTATGGTTCATGCT	8542
2034	eglyAspIleAsnGlnIleProPheIleAsnArgGluLysVal.....	2048
8543	CGGTGATAGTAGACAAATTCATATTATTCACCGGAACGATTTTGGTGT	8592
2049	..PheArgMetAspCysAlaValPheValProLysLysGluSerValVal	2064
8593	CATTGTTGCATGACATAGACCGGTTATA...GAACACACAGCACCGCAT	8639
2065	TyrThrSerLysSerTyrArgCysProLeuAspValCysTyrLeuLeuSe	2081
8640	TACGGTGAAGTCTCTACAGGTGTCTTGGGACATCTGTGAGTGGTGTCTC	8689
2081	rSerMetThrValArgGlyThrGluLysCysTyrProGluLysValValS	2098
8690	T.....AAATTCTATCCGAGACACGTGGCCA	8715
2098	erGlyLysAspLysProValValArg...SerLeuSerLysArgProIle	2113
8716	CTGTCAACACCTGGTAGTATCGGGAAATCGAGTGTGTCTATAGAGATNAT	8765
2114	GlyThrThrAspAspValAlaGluIleAsnAlaAspValTyrLeuCysme	2130


```
/db_xref="GI:7283223"
/translation="GVVRSORIPPRKASLOENLLSYESRNYNFIKTERFVCPENFGKV
MAAAVIERCFMDEVAKIRSDVIALTEANIOKWLDRTPCOVKSVQAEKLPFAVEEQ
ITNKLVMKRAKVKLDDSSLSKHPAALNIMFHKKFIINAIYSPCFDEFKNRVLSSLANG
INVFTEMTAGLAELVIRRIIGDDNLFVGEVDFSKFDKSDQDLFVKFYERALYAEFGF
DAQLDVMGEYERARATLQDLSFVSDGQPRSGGNTWIGNSLVTLGILSLYDDDS
KFELLISGDSLIYSKEIKSINFSSEICLETGFTKMPSPVPYCFKSFVQTKNKT
FVPPYKLLVKLGAPOKLTDELFEFLTSPKDJTOFGDOVLEKLLKLVETKYGFT
SGTTPALCAIHCVRSNYSLEERLFPFIRGWYVVDALKFROLRFANLICERVYDHR
VGYFSFDNPFKPKDASDQGDLLATG"
10835..11746
/function="unknown"
/translation="33-kDa protein"
/citation=[3]
/codon_start=1
/product="p33"
/protein_id="AAB38757.1"
/db_xref="GI:1732496"
/translation="MFAFASENQDILEEKFRRRTCHRYLDDVVVDFTIDIGYDVTVD
RPTVLADLYSLYFLNNDYIGPLAASIMVSPVSGTHKIRAHIDNPQREGNVTYV
KTVKRSFVIRIKAMPASMRGHYSIRLAFSGDVAATVEFICSVFGRFLCCGQTIS
ENLNKVCSSFFFRVSETATNEFSVVTDDVEDVKYVRKRAEGFLHGCADPPYPCRYN
ANLGDGCVGSRTTEEGYRTDIVGNAVTPNPLVNAVSPYVAEYNENARSISLI
RVGCYAVCLVWVSLIMSGLLTILI"
11816..11971
/function="unknown"
/translation="6-kDa protein; a small hydrophobic protein,
homologous to small hydrophobic proteins of other
closteroviruses"
/citation=[3]
/codon_start=1
/product="p6"
/protein_id="AAB38758.1"
/db_xref="GI:1732497"
/translation="MDCVIOGFLFLVGLVAFCAFAELIIIVITRYCTTKPVNRNASP
YGTHTAT"
11977..13761
/function="chaperone"
/translation="65-kDa protein; CTV encoded homolog of HSP70 family
of heat shock proteins"
/citation=[3]
/codon_start=1
/product="p65"
/protein_id="AAB38759.1"
/db_xref="GI:1732498"
/translation="MVLGLDQFGTFTVAMATSELVILKQSNSSYIPTCLFLHADP
NSVGYDAEYLAASGPGSPGKDLRWGCTEKYQSYLHKLSPKYKVIKVEFGIKR
MPVYLSPLNNDGLSVALPSLIALYTKILSDAEVFNVCSTGVCISVPAVNTLQR
AFTQOYSLGIPCVYIINEPFAAYSTLPKLSAHTYLAIVDFGGTDFDVIYSVRL
PTFAVSSGDMNLGGRIDRKLSDNICEADFLPQKELNVSLSKEALSOTDPVKPI
VTHGMBETSIDQTLREITASVFIIRITDILTQVAKSSMPESQSLKLVVVGSSYL
PGLDITLTPVFSVGIYPVEDARTAVARGCALYSECLDGRTKALLIDCIITHLSVTQF
SADSVVAAAGTPIPFCEGRKLTLRKCVSTSNYQARMFEGDYKVFERNERYAASVSL
FTLGVNRRVNDVEMTLVTKVDSMGKVEFLYKGPSELVNVQGTSHYDYVGMPPRTK
LLRLSAKVNAAALVLAITLTREKRERELLRTLDTLLADLRKTSASLDILRYITR
NDIDVSSRMGIVSVKLRASDLERVPL"
13685..15295
/function="unknown"
/translation="61-kDa protein"
/citation=[3]
/codon_start=1
/product="p61"
/protein_id="AAB38760.1"
/db_xref="GI:1732499"
/translation="MSSHVMVWGSIFRKFYGOAIWKVEYLSERTRNFRNERNVSLNHTLSS
GVVVRROSLNAPQGTENELALLYNSVINDFVELTGMPLKSLMTGVEDRKVPDELI
SVDPHEVGCRTFLNDVESYILISRGEDSDGTAVEHTWCLNSCCKLLSSREIDAYTK
VFTKFSGVFGVTTKLETLSYWSLYKHKMKDDDIYNILPMFNCMLKMLVASLGL
FYEKHTNPULTGLMIDFCLEUKVYIYPTFKVNDLNDRFLSKVLPLVLTWQDISEPD
PVDRITLTPDPTDVIDIPKLNIDHTMVVGNQIRHLEHVVEDALDLSQHVRLNL
AANDPLDRWAGMPEVYGIYRCAADRVVVERPALFNPKLLSQDQDGSCLRMSS
VETFLNLVQKVNKDINIRROPFMGRHSKVALRLXNLGLKFPPISSVRLPAHGLLYVD
FYKRIPNDAVADELESRLQRSSVSWCKDRVSITPPPPNRRHRGSSRLIRGRGAK
```

CDS

CDS

CDS

CDS

```
ASSRMSBDVATTGTFNLPHFYHGRJLHAS"
15270..15992
/function="unknown"
/translation="27-kDa protein; a diverged copy of the CTV coat
protein."
/citation=[3]
/codon_start=1
/product="p27"
/protein_id="AAB38761.1"
/db_xref="GI:1732500"
/translation="MAGYTMLPKTKDDKEIDPVSAAVPGKYPDVIEKFTVNRSDALIE
alignment_scores:
Quality: 661.50 Length: 2355
Ratio: 0.610 Gaps: 103
Percent Similarity: 46.030 Percent Identity: 20.425
alignment_block:
US-09-301-906-15 x CTV56902 ..
Align seg 1/1 to: CTV56902 from: 1 to: 19226
298 MetArgGluArgAlaPheValCysThrThrLysGlyTrpCysTrpPheAs 314
:::||||::: ||| |||||::: ||| |||||::: ||| |||||:::
2758 ATTCGTCAATCGCTGAGGCTCTGTTGAGTATCGGAAAGCTTTCTTTAG 2807
314 nAsnGlu.....A 317
:::||||:::
2808 GAGGACGTTAACTTAGGGCGCTCCCTTATGTTTCGAGGTACAGCACC 2857
317 rgleuArgGlyGluIleTyrArgArgCysPheSerSerPheSer 333
|||||::: ::::|::: |||:::
2858 GTCCTCAG...CGTTGTACGGTAAGCGGCT ..... 2886
334 IleGlyPheLeuMetHisLeuGlyPheArgSerLeuLysValIleArgPh 350
:::||||::: ||| |||||:::
2887 .....GTTGAATACGGCTTAGAGGTAT.....TT 2912
350 eAlaGlyThrAsnIleLeuHis.....M 358
:::||||::: |||
2913 TTCAGCAACGCGTCTTACCACCTGTTGCTTTACCGATTCCGCCGAGACCT 2962
358 etProSerLeuAsnGluLysArgThrPheGlyTrpLysGlyGlyAspVal 374
::: |||||::: |||::: |||::: |||:::
2963 TCAAACTCTTAAATAAGTAGTATCATAGATGGGTGGAAGAGGTGCGTTA 3012
375 TyrLeuProAsnValProLysThrAlaIleValIalaGlyAspArgThrAr 391
::: ::::|::: |||||:::
3013 TTAATCACTGACACT .....GATAGATTACG 3038
391 gLeuGlyGlyGluIleLeuAlaSerValAlaAsnAlaLeuAsnGlnGluG 408
|||::: ::::|::: |||::: |||:::
3039 TGGCGTGGGAGTGTCTACGAAAGGTGAAGGGCGTTCCTGCTGAT.... 3084
408 luValTyrSerSerValSerSerIleThrAsnArgLeuValLeuArg 424
||| ::::|::: |||::: |||:::
3085 .....TCTTTACTTGTTCGATCGGTGGAGAAAGATTGTAGAAATTC 3126
425 AspGlnSerAlaLeuLeuSerHisLeuAspThrLysLeuCysAspMetPh 441
:::||||:::
3127 ATCAAACTCTG.....GTTGATACGCA..... 3150
441 eSerGlnArgAspAlaMetIleArgGluLysProSerHisArgCysAspV 458
||| ::::|:::
3151 .....CGGCTAGACCTGGTGTTCAGTACCTT 3178
458 aPheLeuLysProArgGluArgGluLysLeuArgGluLeuPheProGlu 474
:::||||::: |||::: |||::: |||:::
3179 TTCAGATGAGTAGAGAACCAACAGGTGATGCTACTAGAGTTATCCTGAG 3228
475 LeuSerIleGlnPheSerAspSerValArgSerSerHisProPheAlaAs 491
:::||||::: |||::: |||::: |||::: |||:::
3229 TTTAATATAAACTTTACTATTCCTGATACCTCGACACCCCGTGGCTGC 3278
```

CDS

491 nAlaMetArgSerCysPheAsnGlyIlePheSerArgArgCysGlyAsnV 508
::: ||| :::
3279 GGGTCTAGAAGTTGTAGAAATCACTTGTTTCGTAATAATACCGAGTAGC. 3327

508 aICysPhePheAspIleGlySerPheThrTyrlHisValLysAlaGly 524
::: |||||:::
3328 ..GATTATTCGGCAGTAGGTGGTCCCATTTATTTCATTACGTGCCGGT 3375

525 HisValAsnCysHisValCysAsnProValLeuAspValLysAspVally 541
|||||::: |||||:::
3376 CATACAGGAGTTTCACGTITGCAGACCAGTGTTACGATATTAAAGACGCCCA 3425

541 sArgArgIle.....AsnGlulIleLeuPheLeuSert 552
|||||::: |||||:::
3426 CGGTAGAGTCGTTTCGTCATCATCAACTTAGTAAAGTAGCTTTTGATCACA 3475

552 hrlAalGlyAlAspSortyrValSerSerAspLeuLeuThrGluAlaAla 568
|||::: |||||:::
3476 CTGACGGGTCTTAAAGCAGTTT.....GGTGCT 3501

569 SerLysSerValSerTyrCysSerArgGluSerGlnAsnCysAspSerAr 585
::: |||||:::
3502 CTTCAATACGAACACCGTCGTGGTAAACGTTTTTACGGAGGTCTTTTCATGC 3551

585 gAlaaspAlaGlyPheMetValAspValtyrAspIleSerProGlnGlnV 602
|||||::: |||||:::
3552 GTCTGATCGCAGTTTATGTGTGCAAGTTTATGACGTGCCATTTCGCAGAGT 3601

602 aIALaGLuAlaMetAspLysLysGlylAlaLeuValPheAspIleAlaLeu 618
|||||::: |||||:::
3602 TGTGTAGACCTATGTTGACAAAGAAGCTTCGGTGTCTATATGACCGCTG 3651

619 MetPheProValGluLeuLeuTyrglyAsnGlyGluValtyrLeuGluGl 635
::: ||| |||||:::
3652 GTTACGCCGTGGTACCTGTTGGACGCTGTGCAGTCTCTTCATCAAAGA 3701

635 uleuAspThrLeuValLysArgGluGlylAspTyrlLeuAlatyrlAsnValG 652
|||||::: |||||:::
3702 GTTACAGCTGTTCCGGTG....GAGTTGGACCGCTGCCGAANTAGGGTTA 3745

652 lYcLnCys.....GlyGluMetTyrlGluHisSerPheSerAsnValSer 666
|||::: ||| |||||:::
3746 TCTACTGTTTAAACAATTCACGGTATACCCATACCTATTCTACGATATGC 3795

667 GlyPhePheThrPheSerTyrValargThrSerSerGl 679
|||::: |||||:::
3796GAATACATCGGTACTCCTTGTGTTGGTGGATGG 3830

679 yAsnValPhelLysLeuGluTyrlGluGlyTyrlArgCysGlyTyrlHisIsl 696
| |||||::: |||||:::
3831 TTTTCTGTCTACTATAGAGATGGTTTCACTCACAGATCGAGAGTTAACTATT 3880

696 euThrMetCysArgAlaGlnLysSerProGly.....ThrGlu 708
::: |||||::: |||||:::
3881 ATTGTATCACCAAGTCGCGGCTTCTCCGACAATCACGGAGATTAAACGG 3930

709 ValThrTyrArgSerLeu.....ValProse 717
|||::: |||||:::
3931 TTACCGTCTCGTGGTGTGATAGTCACCTTGATCAGGATCAAGATTCCGCG 3980

717 rPheValGlyLysSerLeuValPheIleProValValAlaGlyLysServ 734
::: |||||::: |||||:::
3981 GTACTCTAATAAGACTCGTAGTTGTTTACCA..... 4011

734 aISerPheLysThrIleValLeuAspSerAspPheValAspArgIleTyr 750
|||::: |||||::: |||||:::
4012 ...GGGTGTTATTATTGTTATTAGACCGCTAGTTCGTGCTCGTGTGCTAC 4059

751 SerTyrlAlaLeuAsnThrIleGlyThrlPheGluAsnArgthrPheGluTy 767
|||::: |||||::: |||||:::
4060 GAATACCTCGTTTAAATTAATGTGTGTGGTGAATGCTAAGACTTTTGAATG 4109

[illegible]

4929 TTGGGACGTGGTCAATTTGATTAGATCGTCAACGGGGTGACACGGA 4978
||||:||||| |||:|||||
1030 roValGlyLeuThrSerArgValala.....ValSerLys 1043
||||||| |||: |||
4979 CGAAA.....TTGGAGTCTTTTCGCAATGTTCTTCGGTTAAATCTCTACG 5022
1044 valValAlaGlyThrLys..... 1049
:||||:||||| |||
5023 TTGTGTTCTGGTGTGAAGGACTTCGCAAAAGACTTTGCAAAACATGTGTAG 5072
1050GluPheTrpSer..... 1053
|||||||:|||||
5073 GTTAGTTTTGCCATTTTCTACGGAGTTCTTCAGTAAAGCTAATACTACCCC 5122
1054GluValSer 1056
|||||||:|||||
5123 TGTGTCGGTTTACTTTGGATGTATTCCGACTGCGGACTGATTGAGATTCA 5172
1057 LeuAsnAspPhe.....ThrThrPheValLeuArgAsnLysValle 1070
:||||:||||| |||: |||: |||
5173 GCGCGTAAACATGGGACTAGAGACTTGTGTTTCGATGGCTTCAGCCATTTT 5222
1070 uileGlyIlePheValAlaSerLeuGlyAlaAlaProIleAlaTrpLysT 1087
:||||:||||| |||: |||: |||
5223 TAACGTCGTACTTACCTATCTCTCGGAAGTATTAGCTTTCCGTCGTAT 5272
1087 yArgArgGlyIleAlaAlaAsnAlaArgArgTyrAlaGlySerSerTyr 1103
|||: |||: |||: |||: |||: |||: |||: |||
5273 GCGCGAGAGCTTTTGGCGCCCTCGCTATGCAA...GTGGGATCAAACTAT 5319
1103 1103
5320 TGTTTTAAAGAAAGTTTCGGCAAACTGAAATTTTAGGAGAGACGCTTT 5369
1104 ...GluThrLeuSerSerLeuSerGlnAlaAla..... 1114
:||||:||||| |||: |||: |||
5370 TCGTAGACGGTAAAGTATGCTGTGAGCCCTAAGGCTTTCGACAAACTGT 5419
1115GlyGly 1116
CGATCGACAGTAAGGGTTTAGTGTGTTCCGGTGTATTCTCTGGAGTT 5469
1117 LeuArgGlyLeuThrSerSerThrValSerGlyGlySerLeuValValar 1133
||||||| |||: |||: |||: |||: |||
5470 TTGAGGATAGTATTTCGTCTCTGTTTCCGATGCCGCGCGTGGGTG... 5517
1133 gArgGlyPheSer.....SerAlaValThrValArg.... 1144
||||:||||| |||: |||: |||: |||: |||
5518 ...GGTTATGCAACACAGATTTTCCCTCACTTCCCAACACAGGTGTAG 5563
1145AlaThrValAlaLysArgGlnVal 1152
:||||:||||| |||: |||: |||
5564 CGGAAAAAGTTATCTTTCCGCTGGAGCAGACTAAGGAAGCGTGCATCG 5613
1153 ProLeuAlaLeuLeuSerPheSerThrSerTyrAlaIleSerGlyCysSe 1169
||||||| :||||:||||| |||
5614 CCTTTA.....TACGATTCGTCTAAGAGGATC.....TTAGC 5645
1169 rMetLeuGlyIleTrpAlaHisAlaLeuProArgHisLeuMetPhePhe 1186
:||||| |||: |||: |||: |||: |||
5646 GTCTTTGACTTCGGAGATTTCAGAGTCTTTCCAAAAATTTGGAACTAACTA 5695
1186 heGlyLeuGlyThrLeuLeuGlyAlaAraAlaSerAlaAsnThrTrpLys 1202
|||: |||: |||: |||: |||: |||: |||
5696 ATGGTCTAAGTCTTGTATGGACATGTGAGTCTGTACTCTCAGCGTCT 5745
1203 PheGlyGlyPheSerAsnAsnTrpCysAlaValProGluValValTrpAr 1219
:||||: |||: |||: |||: |||
5746 AATAGTGTGTTAGCAAGGAGCAACGCGTTGCAGCATTTAAAGCGGC 5795
1219 gGlyLysSerValSerSerLeuLeuLeuProIleThrLeuGlyValSerL 1236
||||:||||:||||| |||: |||: |||

5796 GGGGAGTGGTCTTTAGTAAACGTGTGAGCGGTATTTCGTCCTTTCTTTTCGCC 5845
1236 eu..... 1236
||
5846 TAGCTCGGTGTGTTAATGAACAGACAGCGAATAACAATTCTGCTAGTGAC 5895
1237IleIleArgGlyLeuLeuAsnAspThrIleProGlnLeu... 1249
:||||:||||| |||: |||: |||
5896 GAATCTGAGTTGGACGAGTGTGCTTAGTGACAGC...CCCAGTCTTCG 5942
1250AlaTyrValProProValGluGlyArg.... 1258
|||: |||: |||: |||: |||
5943 TGTAAAGGGAAGTGTAAAGCTTTTTCGTCTCATTTCCGAGCGGCTTT 5992
1258 1258
5993 CGGTATGTTGTAGCGCGCTGTAGAAGCCTTGAAGCGTTTAGTCTCTACAC 6042
1259AsnValTyrAspGluThrLeuArgTyrTy 1268
6043 GTCGGTGAAGTGTGTACAGCGCATTTATGGACTCTTTTGTGAGTTATT 6092
1268 rArgAspPheAspTyrAspGluGlyAlaGlyProSerGlyThrGlnHisG 1285
:||||: |||: |||: |||: |||
6093 TGTGTAACCAATGTTCCAGTCCGG.....AAGTGTG 6124
1285 luAlaValProGlyAspAspAsnAspGlySerThrSerValSerSer 1301
|||: |||: |||: |||: |||: |||: |||
6125 AATCTTTTCGTCGCGACAGTACCCGAAAAATTTCTGATTTAGTAAAGGG 6174
1302 TyrAspValValThrAsnValArgAspValGlyIleSerThrAsnGlyL 1318
|||: |||: |||: |||: |||: |||: |||
6175 TATATCACATGCGCAATCTTCATTCGGTCCGTACACGACACTCCAGG 6224
1318 uValThrGlyGluGluThrHisSerProArgSerValGln..... 1332
:||||: |||: |||: |||: |||: |||
6225 GTTATCTGGTGTAGTAGGAGTCTCTTCCAGGTTCTTGGATGGTGGTGC 6274
1333TyrThrTyrValGluGlu...Glu 1339
|||: |||: |||: |||: |||: |||
6275 TCACGGTGTGTAATTCGAATATTTTACCGCGTGGTTGACAGGTTTTCG 6324
1340 valAlaProSerAlaAlaValAlaGluArgGlnGlyAspProSerGlySe 1356
:||||: |||: |||: |||: |||: |||
6325 ATGAGACCTTTATTGAGTCTAGTCCATCGGATG..... 6357
1356 rGlyThrAlaAspAlaMetAlaPheValGluSerValLysLys..... 1370
:||||: |||: |||: |||: |||: |||
6358 .AGGACTGCTCGATAACGTGTTGTAGAAATCTTGAAGAACCTCCAG 6406
1371GlyValAspValPheHisGlnGlnSerSer..... 1381
6407 TTTGTTTAGTTTCGATATACACTTTAATAGAGTGTAGATCTAGGATTTTC 6456
1382GlyGluThrAl 1385
6457 CCTTTGAAGGAATTTCTTTGGGCTTTTGGGTTTGTGTTTGGAAACATTGT 6506
1385 aArgGluValGluValAspGlyLysGlyLeuLeuProGluSerValValG 1402
:||||: |||: |||: |||: |||: |||
6507 TAAGGAACCTTTCTATATGCAATTTACGGAAGTCTCCCGGTTTTCGCTT... 6552
1402 lyGluAlaProThrGlnGluArgGlyAlaAlaAspGlyAsnThrAla 1418
:||||: |||: |||: |||: |||: |||
6553ACGAGGTTGTGTAATAGGCG.....TTC 6576
1419 GlnThrAlaValAsnGluGlyAspArgGluProValGlnSerSerLeuVal 1435
:||||: |||: |||: |||: |||: |||
6577 CGTCGCGGTTTCCAAACGAAGTGAC.....GT 6602
1435 lSerSerProGlnAlaAspIleProLysValThrGlnSerGlu.....V 1450
:||||: |||: |||: |||: |||: |||
6603 CGGTACGCTCAAGTG.....ATCACCGAAGCCGAAACCGAA 6640

1450 aHisAlaGlnLysGluValLysGlnGluValProLeuAlaThrValSer 1466
::: :::: ||||| :::: |::: :|||
6641 TCAGATTTTTCGAGAAAGTCTCGCCGAATAAGGCAGGTATCATGTGCTCT 6990
::: :::: ||||| :::: |::: :|||
1467 GlyAlaThrProIleValAspGluLysProAlaProSerVal 1480
||::: :::: |::: :|||
6691 CTCACATACCCTGCCAGGTCGCCTGATAGTCCTTTTGTTCCACCAAC 6740
||::: :::: |::: :|||
1481ThrThrArgGlyValLysI 1487
.....:~::~:~::~:~::~:~::~:
6741 ATACGGCAGATTGAGGACTTATTAACAGATTCCTCTCACAGTCCCGAT 6790
.....:~::~:~::~:~::~:~::~:
1487 IetIeAsPlysGlyLysAlaValAlaHisAlaGluLysLysGlnVal 1503
:|||| :::::~::~:~::~:~::~:~::~:
6791 TGATAGCTGGACAATGGTTTAGATTGAGTCCCCTGAAGCGCAAGCA 6840
:|||| :::::~::~:~::~:~::~:~::~:
1504 GlnValGluGln.....ProLysGlnArgSe 1512
.....:~::~:~::~:~::~:~::~:
6841 GAAC TAGATTCCGATCGTATAGCCACCACTCAATACCTTTGGATCCGA 6890
.....:~::~:~::~:~::~:~::~:
1512 rLeuThrIleAsnGluGlyLysAlaGlyLysGlnLeuCysMetPheArgT 1529
||::: ||||| :::::~::~:~::~:~::~:~::~:
6891 CTTTGTAGCAAATCGGAGGCTCTGGCCGAAGCGT..... 6924
||::: ||||| :::::~::~:~::~:~::~:~::~:
1529 hrCysSerCysGlyValGlnLeuAspValTyAsnGluAlaThrIleAla 1545
.....:~::~:~::~:~::~:~::~:
6925AACGTTGACGCTCTTCGT.....TTAGGG 6948
.....:~::~:~::~:~::~:~::~:
1546 ThrArgPheSerAsnAlaPheThrPheValAspAsnLeuLysGlyArgSe 1562
:|||| :::::~::~:~::~:~::~:~::~:
6949 CTGAAGTTC.....GTGTGGAGGATGTCGTAATTGTTTAACTTAAG 6992
.....:~::~:~::~:~::~:~::~:
1562 rAlaValPhePheSer...LysLeuGlyGluGlyTyThrTyAsnGlyG 1578
:||::: ||||| :::::~::~:~::~:~::~:~::~:
6993 AGCCGTAAATTTCTCTATGGGTTAGCTCTTCAGTGATATTTAACGCGC 7042
||::: ||||| :::::~::~:~::~:~::~:~::~:
1578 Lyser..... 1579
||::: ||||| :::::~::~:~::~:~::~:~::~:
7043 GTTCAATGGCCAGTTTACGTGGCAGTGTTATTCAGTTTGGCAACTGC 7092
.....:~::~:~::~:~::~:~::~:
1580HisValSerSer...GlyTrpProArgAl 1588
||::: ||||| :::::~::~:~::~:~::~:~::~:
7093 TTAACGTTTGACCTCTTTAGACACAGCTTTGTAGTCTGAGGTCG.....CT 7136
.....:~::~:~::~:~::~:~::~:
1588 alauGluaspIleLeuThrAlaIleLysTyProSerValPheAspHisc 1605
||::: ||||| :::::~::~:~::~:~::~:~::~:
7137 GTCAAGAGCTTTAGCAATACGTGTTTCGTAAACGCPAGAATTTGTACGAAG 7186
||::: ||||| :::::~::~:~::~:~::~:~::~:
1605 ysLeuValGlnLysTyLysMetGlyGlyGlyValProPheHisAlaasp 1621
||::: ||||| :::::~::~:~::~:~::~:~::~:
7187 AACGGTCGAAGCTTCTTCAGTTTTCGCGAATTTGTGAC..... 7227
.....:~::~:~::~:~::~:~::~:
1622 AspGluGluCysTyProSerAspAsnProIleLeuThrValAsnLeuVa 1638
:|||| :::::~::~:~::~:~::~:~::~:
7228 ..GATTTCGGTTACTTATTCGAC.....GTGTTTTCCGGTGTT 7262
.....:~::~:~::~:~::~:~::~:
1638 lGlyLysAlaasnPhe..... 1643
:|||| :::::~::~:~::~:~::~:~::~:
7263 CGAGAGGTCGATTTTTTACTTTAACTATCTATTCCTCTCCCGAAAGGT 7312
.....:~::~:~::~:~::~:~::~:
1644SerThrLysCysArgLysGlyGlyLysValmet 1654
||::: ||||| :::::~::~:~::~:~::~:~::~:
7313 GTCTTACGGGTNACCGTAGTCTCGAGTTACGCGTCGAGCGCCAGTTTCGC 7362
.....:~::~:~::~:~::~:~::~:
1655 ValIleAsnValAlaSerGlyAspTyPheLeuMetProCysGlyPheGl 1671
||::: ||||| :::::~::~:~::~:~::~:~::~:
7363 GTGATTAGG.....GACATGGGCTTGTGGACACGCG..AA 7397
.....:~::~:~::~:~::~:~::~:
1671 nArgThrHisLeuHisSerValAsnSerIleAspGluGlyArgIle.... 1686
||::: ||||| :::::~::~:~::~:~::~:~::~:
7398 GAAGTCTTTAACCCAGGTTGTAACCTCTTTTGAGCTCAATGAAAGTGPCCG 7447
||::: ||||| :::::~::~:~::~:~::~:~::~:

[illegible]

584 eArgAlaAspAlaGlyPheMetValAspValTrAspIleSerProGln 600
::: ||| :::: ||||| ::::: ::::
1863 TGAACAGAGATATAATCATTCAGTTGAGTTTATGACATGACTTTAAGG 1912
601 GlnValAlaGluAlaMetAspLysGlyAlaLeuValPheAspIleAl 617
::::: ||||| :::: ||| ::::: ||||| ::::
1913 GACATGGCAAAAGCTATGTGTCTCATGGAAGCTAGGAAATTTGAGTTTAA 1962
617 aLeuMetPheProValGluLeuLeuTyroGlyAsnGlyGluValTyroLeuG 634
::: :::: ||| ||||| ::::: ||| ||| :::: ||| |||
1963 TTGTATCATACACCTGAGATATTCCTAAGGAGTGTAATGTGGAATGT 2012
634 luGluLeuAspThrLeuValLysArgGluGlyAspTyroLeuAlaTyroAsn 650
||| ||| ||| ||| ||| :::: |||
2013 ATGAAGGCAGACTCAAAAGTACGAGGATGGAGATAATCTAGAAATATAT 2062
651 ValGlyGlnCysGlyMetTyroGluHisSerPheSerAsnValSerG 667
||||| :::: ||||| ::::: ||||| ::::: ||||| ::::
2063 TACGGAGCAACGGAGAGAGCTTTCAGTCATAGTTGTCAGACTCTAAAGA 2112
667 yPhePheThrPheSerTyroValArgThrSerSerGlyAsnValPheLysL 684
::: :::: ||| :::: ||||| ::::: ||||| ::::: ||||| ::::
2113 CATTCATCAGTTCAA...GTCTTCCAATTCGGCGCGCTGTGTTAAGA 2159
684 euGluTyroGluGlyTyroArgCysGlyTyroHisHisLeuThrMetCysArg 700
||| ||| ||| ||| ||| :::: |||
2160 AAACCTTGGAGCAGCTAGAGGACAATTACACTTCTTCCCATATGCATT 2209
701 AlaGlnLys...SerProGlyThrGluValThrTyroArgSerLeuValPr 716
::: :::: ||| :::: ||||| ::::: ||||| ::::: ||||| ::::
2210 TCGGAAAAAATTCACCTCGGAGTGTAATAATTGAAACCTATTACCAGAG 2259
716 oSerPheValGlyLysSerLeuValPheIleProValValAlaGlySerS 733
||| :::: ||| :::: ||||| ::::: ||||| ::::: ||||| ::::
2260 AGTGAATTAGATAAGTCACTCACCTGAGATACCACTTAAGACAGCTTTC 2309
733 erValSerPheLysThrIleValLeu...AspSerAspPheValAspArg 748
::: ||| :::: ||||| ::::: ||||| ::::: ||||| ::::
2310 GGGTGTAAACACATATATATCATAAAAAGAGATAGAGAATTTGTCACTCT 2359
749 IleTyroSerTyroAlaLeuAsnThrIleGlyThrPheGluAsnArgThrPh 765
::: :::: ||||| ::::: ||||| ::::: ||||| ::::: ||||| ::::
2360 ATCATAGATATGTCGGAATACAGGCATCAAAATAAGATGATAAAATGTT 2409
765 eGluTyroAlaValGlyAlaValArgSerGlnLysThrHisValIleThrG 782
||| :::: ||| :::: ||||| ::::: ||||| ::::: ||||| ::::
2410 GGAATGGACCTATTCTCAGTATAGGCTAAGAAACTGTGACTATAAAT 2459
782 lySerArgValValHisSerLysValAspIleSerProAspMetTrp 798
::: ||| :::: ||||| ::::: ||||| ::::: ||||| ::::
2460 CCGGAAGGTAACCTCAAAAGAGACGAGGATTTAGAAAAGAACTAATCCA 2509
799 GlyLeuValValAlaValMetAlaGlnAlaIleLysAspArgAlaLysSe 815
||| :::: ||| :::: ||||| ::::: ||||| ::::: ||||| ::::
2510 GGGTTATAGCCATATCATGTACAGGGGATCAGGGCAAGCAAGAAAGAC 2559
815 rIleArgSerTyroAsnPheIleLysAlaSerGluGlySer... 828
::: :::: ||| :::: ||||| ::::: ||||| ::::: ||||| ::::
2560 TCATTATCTTGCTAAATGTTGTACACCACTCATTTATAACCTTCATAG 2609
829LeuAlaGlyValPheLys 834
||| :::: ||| :::: ||||| ::::: ||||| ::::: ||||| ::::
2610 TTAATATCATCTTTAGTTGATAATGCATTTTGTAGGTGGCACTAACA 2659
835 LeuPhePheGluThrValGlyAspCysPheSerAsnAlaValSerValTy 851
::: ||| :::: ||||| ::::: ||||| ::::: ||||| ::::
2660 TTTATATACGAGTCGTGTGTAGACTGTTTAAATTTTTCACA... 2701
851 rAlaLysAlaMetValHisAspAsnPheAsnValLeuGluThrLeuMetS 868
||| :::: ||| :::: ||||| ::::: ||||| ::::: ||||| ::::
2702AACAGTCACTACATCGATACAAATAGTCA 2729
868 erMetProArgAlaPheIleArgLysValProGlySerValValValThr 884

2730 AC..... 2731
885 IleCysThrSerGlyAlaSerAspArgLeuGluLeuArgGlyAlaPheAs 901
::: :::: ||| :::: ||||| ::::: ||||| ::::: ||||| ::::
2732ACCGAGAGTCGTATAGAA.....GA 2751
901 pIleSerLysGluThrPheGlyArgLysLeuLysAsnSerArgLeuArgV 918
||| :::: ||| :::: ||||| ::::: ||||| ::::: ||||| ::::
2752 CTTAGATAAG..... 2761
918 alPheSerArgAlaIleValGluAspSerIleLysValMetLysAlaMet 934
::: ||| :::: ||| :::: ||||| ::::: ||||| ::::: ||||| ::::
2762TGGTTTGTATTGAAACAAACGCTCACGATA..... 2791
935 LysThrGluAspLysProLeuPro...IleThrGluAspSerValTy 950
||| ||||| :::: ||| ||| ||||| ::::: ||||| ::::: ||||| ::::
2792 ...ACCACGGACGACAGATCAACCTTCTATATTAGAACAAAGTGTAA 2838
950 rAlaPheIleMetGlyAsnValSerAsnValHisCysThrArgAlaGlyL 967
::: ||| :::: ||| :::: ||||| ::::: ||||| ::::: ||||| ::::
2839 ACATTCTTG..... 2848
967 euLeuGlyGlySerLysAlaThrValValSerValSerLysGlyLeu 983
||| :::: ||| :::: ||||| ::::: ||||| ::::: ||||| ::::
2849AGCAAGTATAGTGAAGATGTA 2869
984 ValAlaArgGlyAlaAlaThr.....LysAlaPheSerGlyI 996
||| ||||| :::: ||| :::: ||||| ::::: ||||| ::::: ||||| ::::
2870 TATGAAGGGGTGATCGTAGTTCGAATTAGAACAATTTTCA..... 2914
996 eThrSerPheSerThrGlySerLeuPheTyroAspArgGlyLeuThrG 1013
::: ||| :::: ||| :::: ||||| ::::: ||||| ::::: ||||| ::::
2915AGCG 2918
1013 luAspGluArgLeuAspAlaLeuValArgThrGluAsnAlaIleAsnSer 1029
||| :::: ||| :::: ||||| ::::: ||||| ::::: ||||| ::::
2919 AACAGATCTGGTTGATCATGTTG.....TTGAACCTCT 2950
1030 ProValGlyIleLeuGluThrSerArgValAlaValSerLysValValAl 1046
::: ||| :::: ||| :::: ||||| ::::: ||||| ::::: ||||| ::::
2951 GGA.....GG 2955
1046 aGlyThrLysGlu.....PheTrpSerGluValSerLeuAsnAsp 1060
||| :::: ||| :::: ||||| ::::: ||||| ::::: ||||| ::::
2956 TGGGTCTAAGGAGACGAATGGTTCTTCTTCA.....T 2987
1060 heThrThrPheValLeuArgAsnLysValLeuIleGlyIlePheValAla 1076
::: ||| :::: ||| :::: ||||| ::::: ||||| ::::: ||||| ::::
2988 ATATTTTTCATTAATAGTCAGAACCAAGTGCACCTCTGAGT..... 3025
1077 SerLeuGlyAlaAlaProIleAlaTrpLysTyroArgArgGlyIle..... 1091
::: ||| :::: ||| :::: ||||| ::::: ||||| ::::: ||||| ::::
3026TTAGCCTGGAGATGCAAAATAGAATATTTCAA 3057
1092AlaAlaAsnAlaArgTyroAlaGlySerSer 1103
||| :::: ||| :::: ||||| ::::: ||||| ::::: ||||| ::::
3058 TTTCATAACATCTGTAGACTAAGGGGAAGATTTTGTAAATATATTT 3107
1103 yGluThrLeuSerSerLeuSerGlnAlaAlaGlyGlyLeuArgGly 1119
||| :::: ||| :::: ||||| ::::: ||||| ::::: ||||| ::::
3108 ATGAGATCATATAAATATAATCAACAACCC.....TGGGAAGCG 3148
1120 LeuThrSerSerThrValSerGlyGlySerLeuValValArgArgGlyPh 1136
||| ||||| :::: ||| ||| ||||| ::::: ||||| ::::: ||||| ::::
3149 TTATGTCACTATTTAGTCCAGGATATAAACTGTTATTA..... 3190
1136 eSerSerAlaValThrValThrArgAlaThrValAlaLysArgGlnValP 1153
||| ||||| :::: ||| ||| ||||| ::::: ||||| ::::: ||||| ::::
3191ATAC 3194
1153 roLeuAlaLeuLeuSerPheSerThrSerTyroAlaIleSerGlyCysSer 1169
||| :::: ||| :::: ||||| ::::: ||||| ::::: ||||| ::::

```
3195 CCATGACTTGTGTCGAT...ACTGTAAATTAT..... 3223
1170 MetLeuGlyIleTrpAlaHisAlaLeuProArgHisLeuMetPhePhePh 1186
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
3224 .....TTGACGAGGAGAGTCTAGAGAAATACGTGCTATTT... 3262
1186 cGlyLeuGlyThrLeuLeuGlyAlaArgAlaSerAlaAsnThrTrpLysP 1203
3262 ..... 3262
1203 heGlyGlyPheSerAsnAsnTrpCysAlaValProGluValValTrpArg 1219
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
3263 .....TCTCATGACAGACGCTGATGATGATACATTTTCCC 3298
1220 GlyLysSerValSerSerLeuLeuLeuProIleThrLeuGlyValSerL 1236
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
3299 GGATATGATAGTATTTAGTCAAGAGTCCGGTGATTATGGAGATGAAA 3348
1236 cuIle.lIcArgGlyLeuLeuAsnAspThrIleProGlnLeuAlaThrVa 1252
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
3349 TACTGAATTAGACAG.....TCGCAGTCGTTGTACAG 3380
1252 lProProValGluGlyAcqAsnValTyAspGluThrLeuArgTyTyra 1269
3381 TCATTAAATCAAAAGTAGAAATGTT.....GTTTAGAAATATCAAA 3421
1269 rGaspPheAspTyAspGluGlyAlaGlyProSerGlyThrGlnHisGlu 1285
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
3422 GG.....TCACATCGA 3432
1286 AlaValProGlyAsp.....AspAsnAspGlySerThrSerSerValSe 1300
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
3433 AAAATTCTTCAAGAGATGGGATAAATAGATCAATACATAAGATCTATC 3482
1300 rSerTyAspValValThrAsnValArg AspValGlyIleSerThrAsn 1316
3483 ATGGTTCAAAATCTATTTACACAGCTAGATGTTTAAATTACGATTA 3532
1317 GlyGluValThrGlyGluGluGluThrHisSerProArgSerValGlnTy 1333
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
3533 AATTTTATAACA..... 3544
1333 rThrTyValGluGluGluValAlaProSerAlaAlaValAlaGluArgG 1350
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
3545 .....CACACTGGCTGTGTCATATTATAGCAGTCCCTCA 3578
1350 InGlyAspProSerGlySerGlyThrAlaAspAlaMetAlaPheValGlu 1366
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
3579 CAGGTCAATTCATATAGTGT.....GCGTTAAATTTATGACC 3622
1367 SerVal.....LysLysGlyValAspAspValPheHisGlnGlnSerSe 1381
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
3623 GATGTGGTGATTGAAAGAACTCAGTGAAACACACCACCACT... 3670
1381 rGlyGluThrAlaArgGluValGluValAspGlyLysGlyLeuLeuProG 1398
3670 ..... 3670
1398 luSerValValGlyCluAlaProThrGlnGluArgGlyArgAlaAlaAsp 1414
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
3671 .....AAATTAAATACGA 3682
1415 GlyAsnThrAlaGlnThrAlaValAsnGluGlyAspArgGluProValGl 1431
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
3683 GGCAATTTGACGCAATGCATATACAAATTTGGATTTCATTCATCCAATTTG 3732
1431 nSerSerLeuValSerSerProGlnAlaAspIleProLysValThrGlnS 1448
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
3733 GATCCCTATT.....AGATGGGGGTCAAAGAAATA.....T 3764
1448 erGluValHisAlaGlnLysGluValLysGlnGluValProLeuAlaThr 1464
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
3765 CAAAAATTCACATCAAAAGAAA..... 3787
```

```
1465 ValSerGlyAlaThrProIleValAspGluLysProAlaProSerValTh 1481
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
3788 .....CTTCTAAATTGGCGATCAACCAAGAAAGTGCAC 3819
1481 rThrArgGlyValLysIleIleAspLysGlyLysAlaValAlaHisValA 1498
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
3820 CAACGAGAGATATT.....TGTAAGATTTAGTGCATAAGT 3854
1498 laGluLysLysGlnValGlnValGluProLysGlnArg..... 1511
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
3855 CTTACATCAAAATACATTGATATAAAGAAAGCAGCTTTGGTTGTATGGTTT 3904
1512 SerLeuThrIleAsnGluGlyLysAlaGlyLysGlnLeuCysMetPhear 1528
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
3905 AGTCTTCTAATC.....ATATTTTTCATCC 3930
1528 gThrCysSerCysGlyValGlnLeuAspValTyAsnGluAlaThrIleA 1545
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
3931 AACGTTGGGTTTCTTTATATTGTTTAGCATTTTATCTCGCATTGAGATAA 3980
1545 laThrArgPheSerAsnAlaPheThrPheValAspAsnLeuLysGlyArg 1561
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
3981 AGAGTACTACAAACACGCTAGTCTCTTT...AACACATTAAATGAGT 4027
1562 SerAlaValPhePheSerLysLeuGlyGluGlyTyThrTyTyr..... 1575
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
4028 TATCCACATGTTATTGACAGATTAGAGGTACCTACAATTTCCGCCAATTT 4077
1576 .....AsnGlyGlySerH 1580
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
4078 AAAGAAGCCGTTAGGAGAAATTTTCTGAGAACAAAGTCTCAAGGGTAA 4127
1580 lsValSerSerGlyTrpProArgAlaLeuGluAspIleLeuThrAlaIle 1596
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
4128 AAATTTCAAGAAATACCAGAGAGAGGATGAAGATGTCATGCCACCACATA 4177
1597 LysTyProSerValPheAspHisCysLeuValGlnLysTyThrMetGl 1613
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
4178 GAAGAA.....GTAGAAACACCTAAGAGATC 4203
1613 yGlyGlyValProPheHisAlaAspAsp.....GluGluCysTyTrp 1627
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
4204 AGTAGAAGACCTTCTTCGAAGCAAGATGATATTGAAGATGAATGAATA 4253
1627 roSerAspAsnProIleLeuThrValAsnLeu..... 1637
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
4254 TTAAAGAAGCCCTACTGAGAGTTTAAATTTTCAACATTTGAAGCCACA 4303
1638 .....ValGlyLysAlaAsnPhe..... 1643
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
4304 TCAACAGAGTATGACACTCTTCGTGTAGGAACAGCAAAATTTGCAAAATT 4353
1644 .....SerThrLysCysArgLysGlyLysValMetV 1655
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
4354 CCTAAGATTTTACCCCAAGTCTAAATGT.....T 4382
1655 alIleAsnValAlaSerGlyAspTyThrPheLeuMetProCysGlyPheGln 1671
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
4383 ATGCTAATACATACAGACTGGAGATTTTGATT..... 4411
1672 ArgThrHisLeuHisSerValAsnSerIleAspGluGlyArgIleSerIle 1688
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
4412 AGGGATAGCATAGAGGAATTTTACCATTTAGAAAGAAAGTAAACTGGATAT 4461
1688 uThrPheArgAlaThrArgValPheGlyVal.....GlyA 1701
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
4462 TCAGATCAGTAAATGCAAGGGGTAGTACAGATCATGAATGATCTGCTC 4511
1701 rgMetLeuGlnLeuAlaGlyGlyValSerAspGluLysSerProGlyVal 1717
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
4512 GTATGTTGGATAATGTCACAGCTATGATCGACGACGATCCATGTCAGTT 4561
```


This Page Blank (uspto)